nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

ROLLING CLONES

How genetic evolution alters drug response in cancer cell lines PAGE 325

GENOMICS

A RARE CELL IN THE AIRWAY

Newly identified ionocyte may hold key to cystic fibrosis

PAGES 313, 319 & 377

ORGANIC CHEMISTRY

MOLECULAR SYNTHESIS IN 3D

Cycloaddition reactions expand their range

PAGES 314 & 350

VISION

COLOUR COORDINATION

Songbirds categorize shades of orange and red

PAGES 311 & 365

NATURE.COM/NATURE

16 August 2018

Vol. 560, No. 7718

THIS WEEK

EDITORIALS

WORLD VIEW Funders must stop corporate money from distorting science **p.283**

WILD AT HEART Gene alterations dictate the friendliness of foxes **p.284**

FAST FISH Why an African fish grows to maturity in just two weeks **p.285**

The spectre of smallpox lingers

The last known person to die from the virus was infected 40 years ago. Yet the disease remains a worry, and precautions should continue.

Forty years ago, the UK city of Birmingham was in a panic. A medical photographer called Janet Parker was admitted to hospital on 11 August 1978, her body riddled with lesion and scars. She passed away a month later, the last known person to die from one of the world's deadliest diseases: smallpox. In 1980, the World Health Organization declared that the illness had been eradicated. So it might raise eyebrows that the US Food and Drug Administration (FDA) last month approved the first ever drug to treat it.

The antiviral, called TPOXX (tecovirimat) and made by SIGA Technologies of New York City, inhibits proteins in human cells that allow poxviruses to replicate and spread. TPOXX hasn't actually been tested against smallpox in humans. Rather, the company trialled it in monkeys and rabbits infected with monkeypox and rabbitpox, respectively, and found that it allowed more than 90% of the animals to survive (D. W. Grosenbach *et al. N. Engl. J. Med.* **379**, 44–53; 2018). This was enough for the FDA to approve it for human use, given that clinical trials of such a deadly killer are effectively impossible.

Another antiviral drug, called brincidofovir, isn't far behind. In June, the FDA granted the drug's developer, Chimerix of Durham, North Carolina, a fast-track review by designating it an orphan drug — a therapy for a disease with (all involved surely hope) a small market.

The approvals put the United States on track to fulfil criteria for bioweapons defence drafted in response to the 2001 anthrax attacks that killed five people. In 2009, the US Institute of Medicine recommended that the country stockpile smallpox vaccines and develop at least two antiviral drugs that attack the virus through different molecular mechanisms, to prevent the virus from becoming resistant.

Although it might seem like a waste of resources to develop cures for an extinct disease, smallpox resurgence does remain a realistic threat.

The variola virus that causes the disease officially exists in only two, highly secured places: the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, and its Russian counterpart, VECTOR near Novosibirsk. Every few years, the World Health Assembly debates whether to destroy these last two stocks, and each time, experts postpone the decision. Researchers still use them and, they say, that work should continue, given that we can never be certain the disease is dead.

While the reserves remain, some warn that a disgruntled employee could release the virus — which is what investigators think happened with the anthrax attacks. Or it could be spread through a lab accident. That's what happened to Janet Parker, who worked directly above a university lab that studied the smallpox virus. Investigators later determined that although the researchers had protected themselves, the disease had drifted through air ducts.

The CDC and VECTOR are also unlikely to be the only possible sources. In 2014, the US National Institutes of Health (NIH) discovered live stocks in a storage room on its campus in Bethesda, Maryland. If the venerable and highly regulated NIH could lose track of smallpox, other institutions could have some forgotten vials as well. Even more worrying is the prospect that the virus lives on in freezers in former Soviet states, as US intelligence experts fear, and in countries with illicit bioweapons programmes.

Smallpox might also lie dormant in mummies and corpses of people who died from the disease. As the Arctic permafrost melts, accidental encounters with previously frozen diseases could become more common.

The greatest threat is advances in synthetic biology, which could permit a rogue lab to re-engineer a smallpox virus. In 2016, researchers in Canada announced that they had created horsepox using pieces of DNA ordered from companies. A synthetic smallpox virus could be even more dangerous than the original, because it could be designed to spread more easily or with ways to survive new therapies.

"As permafrost melts, encounters with previously frozen diseases could become more common."

The eradication of smallpox remains one of humanity's most impressive achievements, one that has saved innumerable lives. But it has also left people unvaccinated and vulnerable. New therapies and continued funding for national vaccine stockpiles serve as a crucial insurance policy — just in case. ■

Blurred lines

An in-depth analysis reveals the scale of differences within common cell lines.

Cultures of cell lines are frequently described as the workhorses of medical research — and, like all animals, these ubiquitous biological tools can have their differences. Experiments with cells that work in one lab, or even in one batch, can fail or give different results in another. This can be down to mislabelling, and even contamination (see *Nature* **520**, 264; 2015). In *Nature* this week, scientists highlight another reason: genetic variability in the same cell line (see page 325).

The researchers analysed cell lines from different sources, including the popular MCF7 type used to investigate cancer. They found a staggering amount of variation. Supposedly identical cell lines were actually far from identical — there were differences in their genomes, gene expression, morphology and, importantly, drug sensitivity. This could help to explain why so many experimental results that rely on the standard performance of a specific cell line can't be reproduced.

The cause is genetic drift — the cell lines diverge as they are independently and individually passaged, or subcultured. It isn't an unexpected finding, but has never been shown before on such a scale. It will still probably come as a shock to many researchers: all cell-line strains are not equal. And some are more unequal than others. ■

GEORGE VEKIC

Funders must be wary of industry alliances

Granting bodies have to do more to stop corporate money from distorting science, says **Linda Bauld**.

Do you consider yourself a type A personality — ambitious and impatient? If so, you can thank the tobacco industry for that bit of self-knowledge. For decades, the cigarette companies Philip Morris International and R. J. Reynolds supported and promoted studies linking a driven personality to an increased risk of heart disease. The apparent motivation? To raise questions over smoking as a contributor. Subsequent research by scientists without funding from the tobacco industry did not link type A personalities to higher rates of disease or death.

Copious case studies document how industry influence can muddy research on the health impacts of soda, tobacco, fossil fuels and more, but researchers are largely unaware of this. It is time for research funders to integrate this information and consider vested interests as a force in the complex research system.

Sometimes, partnerships and licensing are essential. Few drugs, for example, would reach patients without the pharmaceutical industry investing in late-stage trials. And plenty of scientific instruments, computing technologies and household products have come from industry-supported researchers. Still, the scientific community must be wary. In my view, industry influence over health research is waxing, not waning.

In June, Francis Collins, director of the US National Institutes of Health (NIH), cancelled a US$100-million study on the effects of moderate drinking. An advisory committee concluded that interactions between government staff and representatives of the alcoholic-beverage industry seemed to be intended to bias the study to find beneficial health effects, and that it was not designed to detect potential harms such as cancer or heart failure.

The NIH is now working to clarify processes for private-sector collaboration, with a plan expected by the end of the year. Funders everywhere should take this as motivation to become more aware of how commercial bias enters the research system — and to learn how to offset it.

Some measures to counter bias are in place. Authors are now routinely asked to disclose conflicts of interest in research papers. And to prevent companies from 'burying' studies that fail to show patient benefit, many countries demand that clinical trials be registered before they start.

These measures are inadequate, and not just because they are enforced imperfectly. Commercial interests can influence research in ways that are largely unmonitored. Industry funding for research on topics such as food, alcohol and tobacco might not directly bias particular studies, but instead encourage a focus on short-term impacts or healthy people, showing less harm. Another tactic is to provide 'unrestricted' grants to researchers whose previous work has aligned with the interests of the industry. In these cases, the researchers may well be acting in good faith, but the industry creates an uneven playing field by providing disproportionate funding to a subset of researchers. Other tactics include possible smokescreens and distractions. For example, in 2017, Philip Morris pledged funds to establish a grant-giving foundation to end smoking, raising all sorts of questions. (The foundation says that, by law, it operates independently of its benefactor.)

Individual funders often have policies or frameworks that are intended to curb industry bias, but these vary widely and are confusing. We need a broad discussion about how industry can support research without skewing it.

Researchers need to reflect on industry influences at several levels, from personal contacts to informational sessions or sponsored research. Informal conversations or interactions at conferences might shape academics' thinking, as might discussions on funding committees, which often include industry members brought in for their perspective. Specifics matter: reformulating or developing a product might require industrial collaboration, but the most peril comes when industries are involved in evaluating the impact on health of that product or making policy recommendations about it.

WE NEED A BROAD DISCUSSION ABOUT HOW INDUSTRY CAN SUPPORT RESEARCH WITHOUT SKEWING IT.

Government and charity funders who advocate industry partnerships should develop and perform structured risk analyses, as should universities and academic researchers who pursue such partnerships. As already required with political donations, public companies should disclose funds given to academics in regulatory filings and name those receiving the most money.

Scientists should get savvier: some ostensibly pro-science initiatives have industry involvement. Vested interests then target regulators by promoting ostensibly pro-science policies that have the effect of discounting or excluding unfavourable results. Industry also provides funds to individuals who make favourable recommendations to advisory panels, as documented in the past decade for dozens of physicians involved in drug approvals.

There is some movement in the right direction. The charity Cancer Research UK, along with several other bodies that award grants for cancer studies, does not fund scientists who are also supported by the tobacco industry. The UK Prevention Research Partnership, a £50-million ($64-million) funding initiative involving a number of government and charity bodies, has developed guidelines on industry partnerships and has indicated that it will support research into commercial influences on health. (It has also funded a £40,000 consortium-development grant from myself and colleagues, held by the University of Stirling, UK.)

Questioning the role of vested interests in research will not be popular or easy, but it is essential. Governments and academics: step up. ■

Linda Bauld *is professor of health policy at the University of Stirling, UK, and deputy director of the UK Centre for Tobacco and Alcohol Studies.*
e-mail: linda.bauld@stir.ac.uk

SEVEN DAYS *The news in brief*

POLICY

US space force

US Vice-President Mike Pence gave details of the White House's plans for its proposed Space Force during a 9 August speech in Washington DC. Pence said the Space Force would be a military branch headed by a new assistant secretary of defence for space, with a suggested budget of US$8 billion over 5 years. The White House wants to establish the Space Force by 2020 to take over warfighting duties in space from the US Air Force, which Pence said would help to counter anti-satellite work by China and other countries, and would strengthen national security. The idea has some support in Congress, which would have to approve the creation of the Space Force. The House of Representatives passed language supporting it in a 2018 authorization bill, but it did not survive the Senate.

EVENTS

Salk settlement

The Salk Institute for Biological Studies has settled two gender-discrimination lawsuits filed last year. Cancer researchers Katherine Jones and Victoria Lundblad have agreed to drop all claims against the institute, located in La Jolla, California, according to a statement released on 7 August by the pair and Salk's president, neuroscientist Rusty Gage. A spokesperson for the Salk Institute declined to comment further on the matter, as did the lawyer who represents Jones and Lundblad. A third gender-discrimination lawsuit against the Salk Institute, filed last year by molecular biologist Beverly Emerson, is still pending. Emerson alleges that systemic bias at the institute hampered her access to research funding, laboratory space and other resources. In January, the Salk Institute declined to renew Emerson's contract; Jones and Lundblad are still on staff. A court in San Diego, California, is scheduled to hold a hearing on Emerson's case on 17 August.

BILL INGALLS/NASA

Spacecraft set for the Sun

On 12 August, NASA launched its US$1.5-billion Parker Solar Probe, a mission that will skim through the Sun's upper atmosphere and 'taste' the source of the solar wind. The spacecraft lifted off from Cape Canaveral, Florida, on a trajectory that will take it past Venus in October, for a gravitational nudge to its orbit, and then past the Sun in November. Over the next 7 years, it will fly close to the Sun 24 times, coming within 6.2 million kilometres of the surface. That will put the craft in the solar corona, where it will sample the magnetic, and other, energies that generate the stream of charged particles known as the solar wind, which drives space weather. It will be, by far, the closest any spacecraft has ever been to the Sun. The mission is named after US astrophysicist Eugene Parker, who proposed the existence of the solar wind and who, now aged 91, attended the launch of the satellite.

Ethics vote

The American Psychological Association (APA) has upheld its ban on psychologists working with the US military at secretive detention facilities. At the APA's annual meeting in San Francisco, California, on 8 August, the organization's governing board voted, 105 to 57, to reject a proposal that would allow psychologists to work with the military at detention sites that have been accused of human-rights violations, even if only to provide care to inmates. The APA first instituted this ethics rule in 2009, after it emerged that psychologists had helped to design torture programmes used at the US detention facility in Guantanamo Bay, Cuba, and other offshore sites.

UNIVERSITIES

Plagiarism rules

India has for the first time introduced regulations for how universities should detect and punish acts of plagiarism. Sanctions for researchers or students caught breaking the rules range from requiring that a manuscript be withdrawn to sacking or expulsion, depending on the extent of the plagiarism. The regulations define plagiarism as "taking someone else's work or idea and passing them as one's own", and will apply to the 867 universities and affiliated institutions that report to the nation's education regulator, the University Grants Commission (UGC). The UGC announced on 3 August that the rules came into effect retroactively from 23 July. Previously, punishments for researchers who were caught plagiarizing were left to the discretion of the institution. The new rules also make it mandatory for institutions to use plagiarism-detection software on students' theses and researchers' manuscripts. Currently, only some universities use detection software. Several high-profile academics in India have been accused of plagiarism over the past decade.

Exam-score scandal

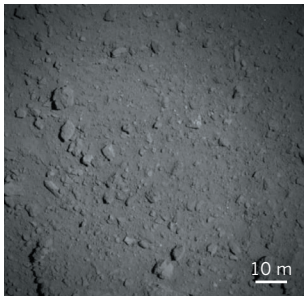
Officials at Tokyo Medical University have confirmed that university staff members have been lowering the scores of female students taking its entrance examination. The announcement follows an article in the Japanese newspaper *Yomiuri Shimbun* that revealed the practice,

prompting an internal university investigation. The investigation found that staff had been reducing female applicants' entry scores by 20 percentage points, with the practice beginning as early as 2006. An anonymous source at the university, quoted in the article, said the practice was designed to decrease the number of women accepted, because of the belief that many leave medicine to look after children. The university investigation found that staff members also lowered the scores of some men who had failed the exam multiple times. At a press conference on 7 August, one of the university's managing directors, Tetsuo Yukioka, apologized for the practice and promised to eradicate it. Three days later, the education ministry launched a survey of the country's 81 medical schools to look for evidence of similar procedures.

SPACE

Asteroid close-up

The Japanese mission Hayabusa2 made its closest approach yet to the asteroid Ryugu on 7 August. Mission control at the Japan Aerospace Exploration Agency first manoeuvred the spacecraft from its hovering position



JAXA/UNIV. AIZU & COLLABORATORS

20 kilometres above the space rock to an altitude of 6 kilometres. Then, it let the probe free-fall to just 851 metres above the surface to measure Ryugu's gravitational pull, and therefore its mass. Ryugu, which is about 1 kilometre in diameter, is a type of asteroid that is common in the Solar System, but which has not been closely studied until now. Newly released images (**pictured**) show a surface strewn with boulders of various sizes. Hayabusa2 was launched at the end of 2014 and arrived at Ryugu in June. It is due to release several landers onto the rock, and the probe itself will also touch down to collect samples, which it will return to Earth in 2020.

FUNDING

Innovation cash

The UK government will invest at least £1 billion (US$1.3 billion) in its network of innovation centres over the next 5 years. The funding, which comes from the research budget and was announced on 10 August, will support six of ten existing centres known as Catapults, which are designed to bring academics and industry together to commercialize research. The six that secured funding focus on areas including gene therapy, semiconductors and renewable energy. The Catapults' governing agency, Innovate UK, says that it will fund the remaining centres in due course. Three of these four facilities are under extra scrutiny after an independent review in 2017 suggested that they might not be delivering value for money.

BUSINESS

Monsanto fined

A California jury has sided with a groundskeeper who says that he developed non-Hodgkin's lymphoma after using the herbicide Roundup, made by the agricultural giant Monsanto. In a decision on 10 August, the jury said that the corporation had failed to warn consumers of the potential dangers of the product, which contains the widely used chemical glyphosate, and the judge ordered Monsanto to pay Dewayne Johnson US$289 million in damages. Glyphosate has become controversial amid arguments over whether it is carcinogenic. In 2015, the World Health Organization's cancer-research arm said that the substance was "probably carcinogenic" to humans, on the basis of studies linking glyphosate exposure to cancer in rodents and humans, but the US Environmental Protection Agency and other bodies say the product is safe, pointing to other studies that find no connection to cancer. Johnson's is the first such case against the company to go to court in the United States; thousands of similar suits are pending. In a statement, Monsanto's vice-president, Scott Partridge, said that the "decision does not change the fact that more than 800 scientific studies and reviews support … the fact that glyphosate does not cause cancer", and that it would appeal against the decision.

CORRECTION

The Seven Days item 'Sterile mosquitoes' (*Nature* **559**, 306–307; 2018) mistakenly said that modified male mosquitoes were released into homes during a trial in Australia. In fact, they were released into neighbourhoods.

TREND WATCH

Tax havens might facilitate deforestation and illegal fishing, according to an analysis published on 13 August.

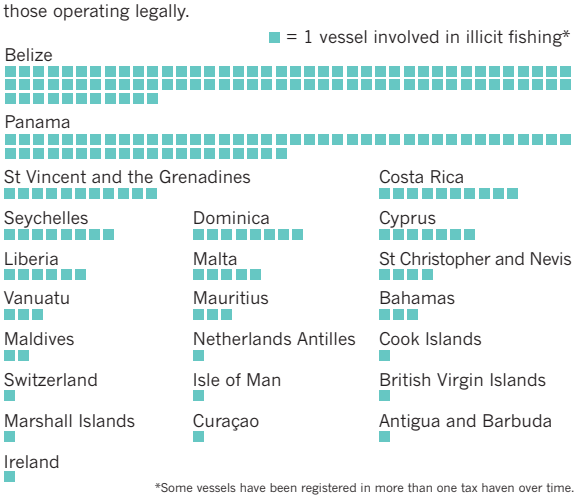
Researchers scoured publicly available databases to find out where ships caught fishing illegally were registered. Of the 209 ships found to be involved in illicit fishing, 70% were flagged under the jurisdiction of tax havens — territories that impose very low taxes and provide secrecy around financial activities. By contrast, only 4.4% of the world's total fishing vessels were registered in these territories.

Most vessels involved in illegal fishing were registered in Belize and Panama, two countries dubbed 'flags of convenience' states. These states impose few repercussions on ship owners who break international law, in addition to providing the financial benefits of a tax haven.

Tax havens might protect unscrupulous fisheries and provide cash for environmentally damaging industries — yet these consequences have been ignored by policymakers, says lead author Victor Galaz, a political scientist at Stockholm University. The new paper provides "the first global assessment to put this issue on the agenda", Galaz says.

FISHING FOR TROUBLE

Vessels cited for illegal, unreported and unregulated fishing are much more likely to be registered in tax-haven territories than are those operating legally.



SOURCE: V. GALAZ *ET AL. NATURE ECOL. EVOL.* HTTP://DOI.ORG/CSZJ (2018)

NEWS IN FOCUS

DAI KUROKAWA/EPA/SHUTTERSTOCK

Movement of refugees and fighters in North Kivu is complicating the response by aid workers to an Ebola outbreak.

PUBLIC HEALTH

War zone complicates roll-out of Ebola vaccine

Latest outbreak comes amid fighting in the eastern Democratic Republic of the Congo.

BY AMY MAXMEN

Aid workers in the Democratic Republic of the Congo began giving an experimental Ebola vaccine to health workers on 8 August — one week after the World Health Organization declared an outbreak of the virus. First responders and public-health staff are scrambling to contain the outbreak while planning how to distribute the vaccine to communities in the middle of a conflict zone.

The virus is spreading in North Kivu and Ituri, in the east of the Democratic Republic of the Congo (DRC). As of 12 August, 57 people had shown symptoms of Ebola — including 41 who had died. But violence perpetuated by more than 100 armed groups fighting over resources in those lush, green provinces has escalated this year ahead of a presidential election scheduled for December. This is the DRC's tenth Ebola outbreak since 1976, but it is the first in this tumultuous eastern region.

"The situation is volatile," says Ibrahima Socé-Fall, director of emergency operations for the World Health Organization (WHO) in Africa, based in Brazzaville in the neighbouring Republic of Congo.

Even so, in addition to dispensing the vaccine, researchers are preparing to give people with Ebola experimental antibody treatments and antiviral drugs.

Socé-Fall says that at least 2,000 doses of the experimental vaccine, called rVSV-ZEBOV, ▶

remain in the country from the most recent Ebola outbreak, which ended in July, and more doses are on the way. The DRC will need a larger vaccine supply now, because the strategy deployed during the previous outbreak will not work for the current one.

During the previous outbreak — which lasted three months — officials vaccinated health workers, people who had had direct contact with someone with Ebola and the contacts of those contacts. But the instability in North Kivu and Ituri has made tracking such connections difficult. In towns where people have been infected but officials can't track down their contacts, workers might vaccinate the entire community instead, says Socé-Fall.

An inability to track these connections worries epidemiologists because people on the move spread the virus. Humanitarian groups estimate that this year, nearly 750,000 people in North Kivu and Ituri have fled from militia fighters. And about one million refugees displaced from their homes by the violence over at least the past decade continue to travel frequently between cities in the region. Some refugees migrate to neighbouring countries such as Uganda, Rwanda and Burundi.

Aid agencies must now consider how to get into these conflict zones to fight Ebola without endangering their staff. Workers might travel with armed security escorts provided by the DRC government for their protection, said Peter Salama, the WHO's head of emergency preparedness and response, at a press briefing on 3 August.

But a key organization fighting Ebola in the area, Médecins Sans Frontières (MSF, also known as Doctors Without Borders), hesitates to use that approach. The group feels that travelling with armed escorts hinders its ability to help people of various political affiliations, says Salha Issoufou, the head of MSF's operation in DRC. So MSF will forgo the escorts.

The next phase of the response by the WHO, the DRC government and aid groups will be to use experimental drugs to treat people who have Ebola. A national review board that evaluates research ethics has approved the use of these treatments, and Steve Ahuka, a virologist at the National Institute for Biomedical Research based in Kinshasa, says that some therapeutics have just arrived in the region.

One treatment is an antibody called mAb114, which was manufactured by the US government. Others include the antiviral drugs Favipiravir (T-705), made by Japanese company Toyama Chemical, and Remdesivir (GS-5734), produced by Gilead, based in Foster City, California.

"Thanks to our experience from the previous outbreak, we are prepared," says Ahuka. ■

Jason Priem (left) and Heather Piwowar co-founded Impactstory, which launched Unpaywall in 2016.

BRIAN GLANZ/OPEN SCIENCE FED./CC BY

DATA ACCESS

The rise and rise of Unpaywall

Non-profit is a gift to many academics — and tie-ins with established scientific search engines could broaden its reach.

BY HOLLY ELSE

After being kicked out of a hotel conference room where they had participated in a three-day, open-science workshop and hackathon, a group of computer scientists simply moved to an adjacent hallway. There, Heather Piwowar, Jason Priem and Cristhian Parra worked all night on software to help academics see how much of their work was freely available on the Internet. They realized how much time had passed only when they noticed hotel staff starting to prepare for breakfast.

That all-nighter, back in 2011, laid the foundation for Unpaywall. This free service locates open-access articles and presents paywalled papers that have been legally archived and are freely available on other websites to users who might otherwise have hit a paywalled version. Since one part of the technology was released in 2016, the service has become indispensable for many researchers. And firms that run established scientific search engines are starting to make good use of Unpaywall.

On 26 July, Elsevier announced plans to integrate Unpaywall into its Scopus database searches, allowing it to deliver millions more free-to-read papers to users than it does currently. Scopus's embrace of Unpaywall, along with similar moves by other search engines, means that much more open-access content is now at researchers' fingertips. These deals are also enabling funders, librarians and others to study open-access publishing trends comprehensively for the first time.

"Unpaywall is a groundbreaking development," says Alberto Martín-Martín, who studies bibliometrics and science communication at the University of Grenada in Spain. "It takes us one step closer to achieving a true open research infrastructure."

After participating in the 2011 hackathon, Piwowar and Priem founded a non-profit organization called Impactstory, in Vancouver, Canada, where they refined Unpaywall. (Parra is now a consultant at the World Bank in Asunción, Paraguay.)

"Unpaywall is a groundbreaking development. It takes us one step closer to a true open research infrastructure."

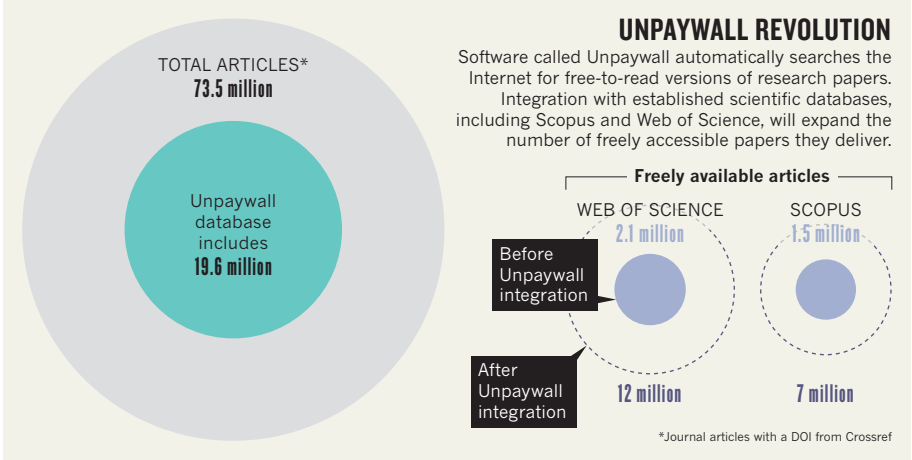
Research by Priem and Piwowar published as a preprint in August 2017 — using Unpaywall, naturally — suggests that almost half of the recent research papers that people search for online are available for free (H. Piwowar *et al.* Preprint at *PeerJ Preprints* https://doi.org/10.7287/peerj.preprints.3119v1; 2017). But, says Priem, "there is a terrific gap between the availability and discoverability" of these papers, and it is this problem that Unpaywall hopes to solve.

Unpaywall consists of a database that includes a list of almost 20 million freely available scholarly articles. Most researchers access it using a free browser plug-in that was released in 2017. In June 2017, Unpaywall was integrated into a popular science search engine called Web of Science, which is operated by Clarivate Analytics. Dimensions, a service run by Digital Science that launched this year, used Unpaywall from the start. These companies, and now Elsevier, pay a subscription fee for a feed of Unpaywall's database that is updated weekly. Impactstory also offers free access to the Unpaywall database (updated twice a year for non-subscribers).

Since its launch, Unpaywall's technology has also been integrated into many university-library discovery systems, so that users can easily find freely available versions of research papers in institutional repositories. These archives, which are operated by universities, funders and others, host the lion's share of articles in Unpaywall's database, but were difficult to search systematically in the past.

Scientists using Scopus can filter their results to find freely available papers, but the database links to only about 1.5 million papers published in fully open-access journals. Once Unpaywall's integration is complete in November 2018, searches carried out on Scopus for free-to-read literature will also find articles on publisher platforms, even if the journal publishes a mix of open-access and paywalled articles.

This will boost the number of freely available articles in Scopus to 7 million, which is still around 13 million articles fewer than are listed in Unpaywall's database (See 'Unpaywall revolution'). This gap exists because Scopus will not initially link to articles posted in repositories.



SOURCE: CROSSREF/UNPAYWALL/IMPACTSTORY/ELSEVIER/CLARIVATE ANALYTICS

NEW FRONTIERS

Large citation databases such as Scopus and Web of Science list the majority of all research articles. By integrating their records with Unpaywall data, researchers can systematically measure the proportion of the literature that is freely available — a feat that wasn't previously possible. The US National Institute of Mental Health (NIMH), which has an overall budget of around US$1.5 billion, is working with Impactstory to develop a bespoke tool that uses Unpaywall. The agency's goal is to determine the extent to which researchers working at NIMH laboratories in Bethesda, Maryland, and nearby Rockville are making their papers, data and source code freely available.

For Priem, making Unpaywall a go-to tool for researchers is just the start. Last month, Impactstory secured an $850,000 grant to create a search engine aimed at non-scientists. It will also use artificial intelligence to summarize journal articles in its database in plain language, so that non-specialists can understand them. "20 million articles are free for everyone to read but might as well be closed if there is no way for any average person to access it," he says. "We're not yet finished." ■

DRUG DEVELOPMENT

Gene-silencing drug approved

US government okays first RNA-interference drug — after a 20-year wait.

BY HEIDI LEDFORD

US regulators have approved the first therapy based on RNA interference (RNAi), a technique that can be used to silence specific genes linked to disease. The drug, patisiran, targets a rare condition that can impair heart and nerve function.

The approval, announced by the US Food and Drug Administration on 10 August, is a landmark for a field that has struggled for nearly two decades to prove its worth in the clinic. Researchers first discovered RNAi 20 years ago (A. Fire *et al. Nature* **391**, 806–811; 1998), sparking hopes of a revolutionary new approach to medicine. Since then, however, a series of setbacks has lessened those expectations.

"This approval is key for the RNAi field," says James Cardia, head of business development at RXi Pharmaceuticals in Marlborough, Massachusetts, which is developing RNAi treatments. "This is transformational."

Patisiran works by silencing the gene that underlies a rare disease called hereditary transthyretin amyloidosis. In that illness, mutated forms of the protein transthyretin accumulate in the body, sometimes impairing heart and nerve function.

The drug's approval means that pharmacology textbooks will need to be rewritten, says Ricardo Titze-de-Almeida, who studies RNAi at the University of Brasilia. "We are inaugurating a new pharmacological group," he says. "We will have many more such drugs in the coming years."

This was the hope when Alnylam, the company in Cambridge, Massachusetts, that developed patisiran, launched in 2002. Four years later, the Nobel Prize in Physiology or Medicine was awarded to two RNAi pioneers: Andrew Fire of Stanford University School of Medicine in California and Craig Mello of the University of Massachusetts Medical School in Worcester.

But to make RNAi into medicine, developers would first need to determine how to deliver delicate molecules of RNA safely to their target organs. They needed a way to shield the RNA from degradation in the bloodstream, prevent it from being filtered out by the kidneys, and allow it to exit blood vessels and spread through tissues. "That proved to be a substantially harder problem than we anticipated," says Douglas Fambrough, chief executive of Dicerna, an RNAi-focused company in Cambridge, Massachusetts.

As researchers grappled with the delivery puzzle, investors began to lose confidence. In 2008, analyst Edward Tenthoff of investment bank Piper Jaffray in New York City advised his clients to stop buying Alnylam stock. "We saw the promise in the technology, but the delivery was lacking," he says. ▶

By 2010, large pharmaceutical companies were also losing their appetite for RNAi, severing collaborations and ending internal research programmes. "By and large, big pharma left RNAi for dead," says Fambrough. Safety concerns dealt the field another blow in 2016, when Alnylam abandoned one of its leading RNAi programmes after finding a possible link to patient deaths in a clinical trial (see 'Ups and downs').

But gradually, some RNAi companies began to iron out the kinks in their delivery systems. Alnylam experimented with a number of delivery routes and target organs, encasing some of its RNA molecules in fatty nanoparticles or chemically modifying the RNAs to help them survive the perilous journey through the bloodstream.

RNAs protected in this way and injected into the bloodstream tended to accumulate in the kidneys and liver. This led the company to look at transthyretin, which is produced mainly in the liver. In a clinical trial in 225 people with hereditary transthyretin amyloidosis who showed signs of nerve damage, average walking speed significantly improved in those who received the treatment (D. Adams *et al. N. Engl. J. Med.* **379**, 11–21; 2018). Walking speed declined in the placebo group.

UPS AND DOWNS

The biotech firm Alnylam faced several setbacks before winning US government approval for its first RNA-interference drug.

SOURCE: NASDAQ

In the future, Alnylam and others will be able to move beyond the liver, says company co-founder Thomas Tuschl, a biochemist at Rockefeller University in New York City. Quark Pharmaceuticals of Fremont, California is testing RNAi therapies that target proteins in the kidneys and the eye. Alnylam is developing ways to target the brain and spinal cord, and Arrowhead Pharmaceuticals of Pasadena, California, is working on an inhalable RNAi treatment for cystic fibrosis.

"I've never been more optimistic about the future of RNAi," says Fambrough. "All of those tear-your-hair-out days were worth it to get to today."

Advances in RNA delivery might also benefit researchers who are developing gene-editing therapies based on the popular technique CRISPR–Cas9. That system uses a DNA-cutting protein called Cas9, which is guided to the desired site in the genome by an RNA molecule.

Like RNAi before it, CRISPR–Cas9 has become a common tool in genetics laboratories. But it might still face a difficult and lengthy path to the clinic. Much like ordinary drugs, RNAi therapies will break down over time; a gene edit, however, is intended to be permanent, which amplifies safety concerns.

"I hope they can do it more quickly than we did it, but I would not expect it to be so smooth," says Fambrough. "I wish them the best of luck." ■

POLICY

Outrage over changes to EPA chemical assessments

Critics say US environment agency's revisions favour industry over academic research.

BY JEFF TOLLEFSON

The US Environmental Protection Agency is making major changes to the way in which it evaluates chemicals for environmental and public-health effects. The latest push includes changes to chemical-safety guidelines that place greater weight on industry-sponsored research, among other things, and is a part of efforts by US President Donald Trump's administration to reshape how the agency uses science to make decisions.

The Environmental Protection Agency (EPA) issued its chemical-assessment guidance in May, and is soliciting public comments until 16 August. The guidance contains changes dictating the kind of data that studies must include in order to be considered in the EPA's decision-making process. Researchers and environmental and public-health advocates say that the guidelines provide a non-peer-reviewed alternative to the EPA's main system for conducting chemical reviews and calculating acceptable exposure limits. The agency is required by law to do these evaluations, but the guidance defines how officials conduct them. At stake are tens of thousands of chemicals destined for public use and governed by the 1976 Toxic Substances Control Act (TSCA).

The guidance dovetails with a rule proposed in April by then-EPA administrator Scott Pruitt, which, if finalized and implemented, would reduce the role of published scientific studies in decision-making across the agency. The changes also coincide with attacks on the EPA's core chemical-assessment programme, known as the Integrated Risk Information System (IRIS), by industry and Republican politicians over the past year.

In a statement to *Nature*, the EPA says the changes are meant to provide clear criteria to help determine the quality of the research used to evaluate chemicals — and that the guidance is a work in progress that can be revised in response to new information. But scientists say the process laid out by the EPA is at odds with established, peer-reviewed procedures for such assessments.

Jennifer Sass, a senior scientist at the Natural Resources Defense Council, an advocacy group based in New York City, suspects that the goals are to promote science from industry and change the calculations that the EPA uses to develop regulations and estimate safe exposure limits for chemicals.

MEETING THE REQUIREMENTS

The guidelines introduce many data reporting requirements — including statistical analyses that measure whether a study correctly identifies the presence of an effect — that are standard for industry-funded research. But because such criteria vary among peer-reviewed journals, many academic studies would be disqualified, says Tracey Woodruff, who led the development of a chemical-evaluation process at the University of California, San Francisco. "Only industry studies will survive."

The changes represent a major shift because they create a new system for chemical-risk assessments under TSCA. Unlike

IRIS, the process introduced by the Trump administration has not been peer reviewed, and yet it would allow agency officials to circumvent IRIS evaluations. Under former president Barack Obama, the EPA would have used IRIS to perform these reviews when considering regulations under TSCA.

The IRIS programme dates back to 1985, but under the Obama administration, the EPA modernized and standardized its chemical-evaluation procedures to improve transparency and confidence in its health assessments. Woodruff says that the IRIS process is solid and that bypassing it would be a mistake.

"The TSCA office is deciding to ditch all of the experts and empirical methods that have been developed over the last 30 years for a method that appears to be based on their whim and personal opinion," she says.

But the EPA insists that the review process used in these chemical evaluations is intended to "comprehensively capture all available science".

Politicians in the US House of Representatives have also hammered IRIS, holding hearings questioning the quality and validity of the programme's assessments. The political manoeuvring parallels efforts from industry to bypass scientific reviews of certain chemicals.

One plant in LaPlace, Louisiana, makes the chemical chloroprene for the Tokyo-based company Denka. Chloroprene is used to make neoprene, a synthetic rubber integral to products such as wetsuits. A 2010 IRIS evaluation and subsequent government studies suggested that chloroprene exposure levels in LaPlace were high enough to increase cancer risk in some areas of the city. Denka challenged that ruling last year, arguing that the assessment was incorrect. The company lost its challenge in January but has since appealed against that ruling. A panel appointed by the EPA leadership will now consider the appeal.

Denka has argued to its political allies that reducing chloroprene emissions would be too expensive, says Karl Brooks, a former EPA official who last year served as a consultant in a lawsuit filed by LaPlace residents against Denka. That's a potentially dangerous development, he says, because IRIS assessments are meant to focus on the health effects of chemicals — not the economic challenges that a company might face as a result of the core science.

Researchers fear that the chloroprene case represents yet another strategy for companies seeking relief from the burdens of regulations: challenge the science and, when that fails, appeal to friendly politicians and political appointees. ■

WILLIAM WIDMER/REDUX/EYEVINE

A chemical produced in a US-based plant is part of a challenge to a government safety programme.

PARTICLE PHYSICS

LHC teams turn to brute-force hunt

World's most-powerful particle collider is using a fresh approach to find evidence of 'new' physics.

BY DAVIDE CASTELVECCHI

A once-controversial approach to particle physics could soon have an increased role at the Large Hadron Collider (LHC). The LHC's major ATLAS experiment has officially thrown its weight behind the method — an alternative way to hunt through the reams of data created by the machine — as the collider's best hope for detecting behaviour that goes beyond the standard model of particle physics, because conventional techniques have so far come up empty-handed.

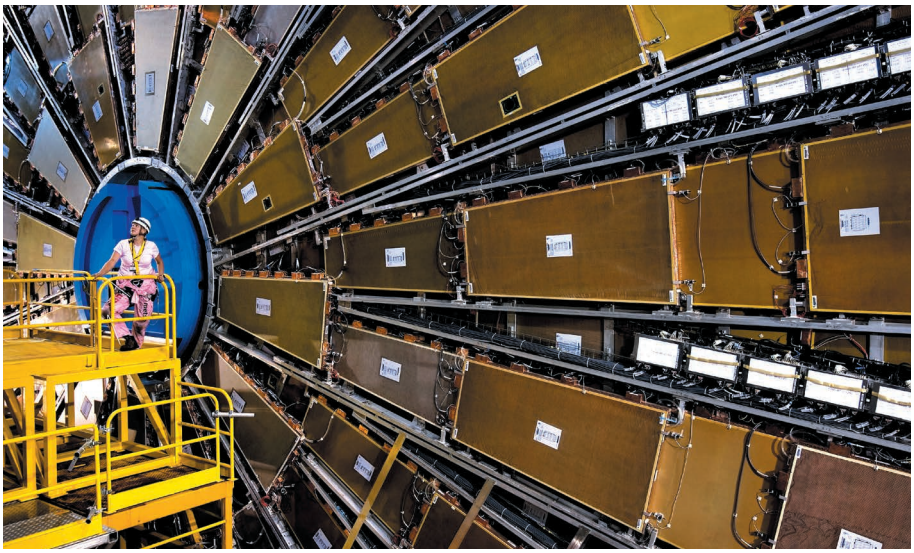
So far, almost all studies at the LHC — at CERN, Europe's particle-physics laboratory near Geneva, Switzerland — have involved 'targeted searches' for signatures of favoured theories. The ATLAS collaboration now describes its first all-out 'general' search of the detector's data — a kind of brute-force approach — in a preprint posted last month and submitted to *European Physics Journal C* (ATLAS Collaboration. Preprint at https://arxiv.org/abs/1807.07447v1; 2018). Another major LHC experiment, CMS, is working on a similar project.

"My goal is to try to come up with a really new way to look for new physics" — one driven by the data rather than theory, says Sascha Caron of Radboud University Nijmegen in the Netherlands, who has led the push for the approach at ATLAS. General searches are to the targeted ones what spell-checking an entire text is to searching for a particular word. These broad searches could realize their full potential soon, when combined with increasingly sophisticated artificial-intelligence (AI) methods.

LHC researchers hope that the methods will lead them to their next big discovery — something that hasn't happened since the detection of the Higgs boson in 2012, which put in place the final piece of the standard model. The model describes all known subatomic particles, but physicists suspect that there is more to the story — the theory doesn't account for dark matter, for instance. But big experiments such as the LHC have yet to find evidence for this behaviour. That means it's important to try new things, including general searches, says Gian Giudice, who heads CERN's theory department and is not involved in any of the experiments. "This is the right approach, at this point."

COLLISION COURSE

The LHC smashes together millions of protons per second at colossal energies to produce a profusion of decay particles, which are recorded by detectors such as ATLAS and CMS. Many different types of particle interaction can produce the same debris. For example, the decay of ▶



STEFANO DAL POZZOLO/CONTRASTO/EYEVINE

The ATLAS detector at the Large Hadron Collider near Geneva, Switzerland.

a Higgs might produce a pair of photons, but so do other processes. So, to search for the Higgs, physicists first ran simulations to predict how many of those 'impostor' pairs to expect. They then counted all photon pairs recorded in the detector and compared them to their simulations. The difference — a slight excess of photon pairs within a narrow range of energies — was evidence that the Higgs existed.

ATLAS and CMS have run hundreds of these targeted searches to look for particles not in the standard model, but the searches have come up empty so far. This leaves open the possibility that there are exotic particles that produce signatures no one has thought of — something that general searches have a better chance of finding.

Whereas targeted searches typically look at only a handful of the many types of decay product, the latest study looked at more than 700 types at once. The study analysed data collected in 2015, the first year after an LHC upgrade raised the energy of proton collisions in the collider from 8 teraelectronvolts (TeV) to 13 TeV. At CMS, Meyer and a few collaborators have conducted a proof-of-principle study, which hasn't been published, on a smaller set of data from the 8 TeV run. Neither experiment has found significant deviations so far. This was not surprising, the teams say, because the data sets were relatively small. Both ATLAS and CMS are now searching a larger trove of data.

The approach "has clear advantages, but also clear shortcomings", says Markus Klute, a physicist at the Massachusetts Institute of Technology in Cambridge who is part of CMS and has worked on general searches for previous experiments. One limitation is statistical power. If a targeted search finds a positive result, there are standard procedures for calculating its significance; when casting a wide net, however, some false positives are bound to arise, one reason that general searches had not been favoured in the past. But the teams say they have put a lot of work into making their methods more solid.

Proponents of this approach hope to use machine learning to find patterns in the data without any theoretical bias. "We want to reverse the strategy — let the data tell us where to look next," Caron says. ■

Philadelphia in Pennsylvania recorded more than 1,200 overdose deaths in 2017.

ADAPTED FROM JEROME SESSINI/MAGNUM

Model citizens

HOW DIGITAL DRUG DEALERS AND VIRTUAL USERS ARE PROVIDING CLUES TO HELP STOP THE US OPIOID EPIDEMIC.

By Sara Reardon

With the tip of her syringe, Brandi pokes at a grey lump of heroin in a spoon. It's a new variety of the drug that has shown up on the market in the past few days, and Brandi likes it. "I feel this more, I feel more of the pain resistance," she says.

Once it has dissolved into a liquid, she injects it into her arm, then uses a fresh needle to inject the skinny arm of another woman. "She does it better than the hospital," the woman comments.

"I'll help anybody who needs it," Brandi explains to public-health researcher Daniel Ciccarone of the University of California, San Francisco, who has been filming the entire process.

Ciccarone's team has embedded with Brandi — whose name has been changed for this story — in Charleston, West Virginia, documenting her interactions without judgement or interference. Later, the group will analyse this video, in addition to half a dozen other videos of drug users from across the city, logging details big and small. Brandi does not heat the solution on the spoon, for instance, and that may increase the likelihood of spreading viruses such as HIV. And tests reveal that what she's taking has been laced with fentanyl, a synthetic drug up to 50 times more powerful than heroin.

The researchers will plug these data into powerful computer simulations of Charleston, populated by thousands of virtual Brandis — heroin users and dealers going about their daily routines. They will watch these digital agents buy more heroin as their tolerance increases, form networks with sellers and users and, in some cases, accidentally overdose.

Ciccarone's is one of several groups using agent-based models to understand what is driving the US opioid epidemic — the dramatic rise over the past two decades in the use of opioids, including prescription pain medications and illegal drugs such as heroin. By studying the motivations and practices of real drug dealers and users, the researchers hope to build agents whose behaviour in the virtual world mimics that in real life.

Agent-based models promise to provide a more granular view of the opioid crisis than standard modelling, which is based on average populations, and to capture some of the complexity of the driving forces. This could prove important for demonstrating the effects of opening or closing methadone clinics or needle exchanges. The models allow scientists to compare interventions at almost no cost and could help policymakers to decide how to proceed in the real world. "It's a very classic and useful way to try and see where is the best place to deploy an intervention to have the biggest effect," says John Brooks, a medical adviser for the division of HIV/AIDS prevention at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia.

Although such simulations have long been used to model disease outbreaks and have, in some instances, guided public policy, their track record with more complex social behaviour such as drug use is limited,

largely owing to sparse data and the breadth of parameters to consider.

Still, scientists hope that agent-based models can lay out scenarios for decision-makers, who are often driven more by politics than data. "The barriers are not scientific or medical," Ciccarone says. "You can throw $1 billion at West Virginia, and they may or may not know how to use it well." These virtual worlds can add clarity, says Joshua Epstein, director of the Agent-Based Modeling Lab at New York University. "You can literally watch the thing unfold before your eyes," he says.

THE DIFFERENCE IN THE DETAILS

The US opioid crisis is estimated to kill 115 people per day through overdoses, and has run up US$1 trillion in health-care costs and lost productivity since 2001. It is not the first addiction crisis that the United States has faced, nor is it the most severe. Alcohol use causes many more deaths, and the rate of cocaine overdose among African Americans is similar to the rate of opioid overdose in white Americans[1].

But the opioid crisis does have some different driving factors, including the prevalence of prescription drugs, which many have used on the way to abusing illegal drugs, and the introduction of fentanyl, which is often used to boost the potency of heroin and is responsible for a large share of overdose deaths. The epidemic has also hit hard in rural settings, where services and infrastructure for dealing with addiction are scarce. "The demographic now encompasses a population that in the past has not been so affected," says Nora Volkow, director of the US National Institute on Drug Abuse in Bethesda, Maryland.

As a result, researchers are coming up with fresh ways of thinking about the crisis. It bears similarities to a disease epidemic, for example, in the way it spreads through networks based on personal relationships and physical proximity, says Georgiy Bobashev, a data scientist at the non-profit research institute RTI International in Research Triangle Park, North Carolina. "Nobody is born an addict. Somebody has to teach you how to smoke or how to inject."

These personal networks can be replicated using agent-based modelling. Unlike other types of model, which may rely on average characteristics or relationships between homogeneous groups to inform algorithms, agent-based models allow researchers to see subtle connections between people. "That's useful because drug use and overdose is inherently personal," says epidemiologist Brandon Marshall of Brown University in Providence, Rhode Island. Factors such as job loss, mental health or genetics can influence how likely a person is to begin using drugs or become addicted, but those factors might fade into the averages if researchers looked at a population as a whole.

To create an agent-based model, researchers first 'build' a virtual town or region, sometimes based on a real place, including buildings such as schools and food shops. They then populate it with agents, using census data to give each one its own characteristics, such as age, race and income, and to distribute the agents throughout the virtual town.

The agents are autonomous but operate within pre-programmed routines — going to work five times a week, for instance. Some behaviours may be more random, such as a 5% chance per day of skipping work, or a 50% chance of meeting a certain person in the agent's network. Once the system is as realistic as possible, the researchers introduce a variable such as a flu virus, with a rate and pattern of spread based on its real-life characteristics. They then run the simulation to test how the agents' behaviour shifts when a school is closed or a vaccination campaign is started, repeating it thousands of times to determine the likelihood of different outcomes.

In 2015, data from an agent-based model[2] developed at the University of Pittsburgh in Pennsylvania helped California state senator Richard Pan to gain support for a bill on mandatory vaccination in his state. Pan used the simulation to demonstrate to his fellow senators how measles outbreaks could unfold in their home districts. "It certainly made an impact on them," Pan says. "Instead of just describing it in more abstract terms, [the model] can make it more concrete." The bill ultimately passed, and immunization rates increased.

As computers have improved, researchers have begun adapting agent-based models to look at sociological and behavioural trends that require more computing power because of the number of variables they contain. Some groups use the technique for crisis modelling, and Australia has begun intervention studies for child obesity on the basis of the findings of an agent-based model.

In response to the opioid epidemic, Bobashev's group has constructed Pain Town — a generic city complete with 10,000 people suffering from chronic pain, 70 drug dealers, 30 doctors, 10 emergency rooms and 10 pharmacies. The researchers run the model over five simulated years, recording how the situation changes each virtual day.

During this time, the patients' drug tolerance increases, leading them to find different ways of acquiring drugs. Their behaviour is driven by variables such as the chance that a doctor will increase their prescription, or the likelihood that a dealer will have enough heroin. At a certain threshold, patients become addicted or more likely to overdose. Bobashev's early data suggest, for example, that requiring doctors to track patients' medication history can be effective over the long term, but not immediately[3].

The model contains many assumptions and simplifications, Bobashev says. For example, it doesn't capture the fact that the rate at which people develop tolerance and addiction can depend on factors such as genetics, and that whether a person switches from prescription drugs to heroin can depend on the relative availability of the two drugs.

But researchers can adjust models such as Pain Town to test various interventions, such as increasing access to emergency rooms, arresting a dealer or equipping police with naloxone (a drug that reverses opioid overdoses), to see how the system reacts and whether it affects the number of deaths over time. And as models become more sophisticated, the researchers may be able to incorporate more factors, such as people who are not taking pain medications but are susceptible to trying opioids for the first time.

Models can also be useful for understanding why individual places or situations may differ, says Christopher Barrett, a computer scientist at Virginia Tech in Blacksburg. For instance, heroin and fentanyl might be easier to come by in cities near ports, whereas doctors may be the main source of opioids in a suburban or rural setting. Interventions focused on prescribing practices, therefore, would have different effects in each case.

> "We're trying to enter their world as interlopers to see how they see their life."

Such models can also reveal feedback loops, such as the link between economic downturns and opioid use. Some epidemiological studies have suggested that factors such as unemployment tend to predict suicide and addiction, especially in white male populations. And addiction can lead to further job loss and lower productivity, harming the economy. Agent-based models could investigate loops such as this, providing ideas for how to mitigate the effects, Barrett says.

In May, Bobashev and Ciccarone presented results from one of their agent-based models at a meeting of the International Society for the Study of Drug Policy in Vancouver, Canada. Their findings suggested that the increased prevalence of white-powder heroin — a newer form of the drug in the United States — may increase the risk of HIV spreading among injection drug users. The reason, also supported by the model, is that unlike black-tar heroin, users don't need to heat the drug to dissolve it — and heating kills the virus.

Bobashev and Ciccarone are working on models of how younger heroin users begin using the drug. Unlike older users, who experienced the rise of the HIV epidemic in the 1980s, newer users may be less likely to adopt safe practices. The models suggest that the United States may see more localized HIV outbreaks, similar to the recent outbreak in Scott County, Indiana. That region experienced 181 new HIV cases

Youngstown in Ohio has high rates of unemployment and opioid addiction.

SPENCER PLATT/GETTY

between November 2014 and November 2015, compared with fewer than 5 cases per year previously. Opioid use is thought to be the cause. Agent-based models might help to stem future outbreaks by guiding surveillance priorities.

One of the most sophisticated agent-based models is the University of Pittsburgh's system, known as FRED (A Framework for Reconstructing Epidemiological Dynamics). It fits population census data to maps of geographical regions around the country, allowing researchers to track virtual individuals in the area in a realistic way. It was data from these models that helped to convince Pan and his fellow state senators to pass legislation on mandatory vaccination. The FRED team is now beginning to use the system for opioid modelling, training it on historical trends. Pan, who is also a physician, says he is intrigued by the prospect. "If there's a way to actually model in different communities which factors would have the biggest impact, that would be helpful," he says.

DATA DROUGHT

The models face numerous challenges before they will be ready for widespread adoption, primarily data gaps. Marshall says that researchers struggle to get access to data on opioid prescriptions that are held by manufacturers, pharmacies and law-enforcement agencies. It is also difficult to obtain government information on drug cartels and the type and rate of drugs flowing into the country. Other data simply do not exist in usable form: agencies may record deaths due to drug overdose, for instance, but fail to specify which drug was responsible.

Observing drug users such as Brandi can provide certain types of information more quickly and accurately. "Drug users know their chemicals intimately," Ciccarone says.

Lee Hoffer is a cultural anthropologist at Case Western Reserve University in Cleveland, Ohio, who studies heroin markets and collaborates with Bobashev. He says the ethnographic data that his group and others are collecting could help to fill some of the information gaps: "We're trying to enter their world as interlopers to see how they see their life." After an initial awkward period, he says, drug users tend to become more honest with the researchers, telling them crucial information such as how they form networks with dealers and the cost of drugs.

Understanding the psychology of drug users is also crucial, says Epstein. Most decision-making models assume rational behaviours. In reality, emotions, misinformation and irrational calculations play a major part. "When you put them together you get collections of dynamics that are very dysfunctional."

Epidemiological data may soon be available to buttress the models. The CDC and the US National Institute on Drug Abuse have started several major surveys of drug-use patterns. A number of states have also begun collecting epidemiological information on trends in overdose and addiction. And research groups such as the University of Pittsburgh team are working with multiple health agencies to collate their findings in a single database, which can inform FRED and other models.

But no matter how advanced the models become, implementing interventions based on their findings is an enormous challenge. Models may reveal socioeconomic contributors that cannot be easily addressed by policies, and politics can stand in the way of proven solutions. In April, Ciccarone had to cancel his work in Charleston, at least for the time being, after a needle-exchange clinic with which he had been collaborating closed owing to political pressure. "They were seeing 300 people on a Wednesday afternoon because there's a lot of need," he says. "It's a huge loss."

Increasing work is being done to determine the relative impact of interventions. The US National Institutes of Health (NIH) in April announced $96 million for a programme that will partner with healthcare systems and local governments to carry out evidence-based public-health interventions in different locations, evaluating them as they go. "This is the first time this [has been] done for a particular substance-abuse disorder," says Volkow. The NIH is now asking researchers who want to apply for these funds to justify the size and scope of their proposed studies with data from models, including agent-based models.

But these studies will take many years to complete. And Bobashev says that society cannot afford to wait for the science to be perfect. "By the time this data is collected, tens of thousands, if not hundreds of thousands, more deaths will have occurred." ■

Sara Reardon *writes for* Nature *from Washington DC.*

1. Seth, P., Scholl, L., Rudd, R. A. & Bacon, S. *MMWR Morb. Mortal. Wkly Rep.* **67**, 349–358 (2018).
2. Grefenstette, J. J. *et al. BMC Public Health* **13**, 940 (2013).
3. Bobashev, G., Goree, S., Frank, J. & Zule, W. in *Social, Cultural, and Behavioral Modeling* (eds Thomson, R. *et al.*) 274–285 (Springer, 2018).

COMMENT

ALEX MUSTARD/NPL

Fish and other marine life are affected by ocean weather: drastic variations in temperature, pH, oxygen and salinity that are in turn influenced by climate change.

Biologists ignore ocean weather at their peril

Ecologists must understand how marine life responds to changing local conditions, rather than to overall global temperature rise, say **Amanda E. Bates** and 16 colleagues.

The ocean can turn on a dime. Temperature, pH, oxygen levels and salinity can vary drastically — across distances of centimetres and within time frames of minutes[1–3]. That's the latest view being revealed by measurements from thousands of instruments anchored to shores or attached to floats, ocean gliders and ships.

Yet many people think of oceans as a relatively constant environment. That idea might have been hatched when researchers on the HMS *Challenger* expedition of 1872–76 tracked water temperature and currents and lowered weights to gauge depth at thousands of sites across the world's seas[4]. The global picture that emerged after averaging these data was one of stability, in which any variability had been lost. Certainly, that picture was reinforced by twentieth-century images of Earth from space, showing the world's ocean as a uniform deep blue[5].

Most biologists and ecologists trying to understand how ocean biodiversity is affected by climate change focus on large-scale averages across space and time. They try to predict, for instance, how a mean global temperature rise of 2 °C could affect marine ▶

life such as bacteria, phytoplankton, fish and other creatures. To do this, they use projected changes in the mean temperature of the ocean. These are based on estimates from satellites, which measure the temperature of only the top few millimetres of seawater.

But organisms experience and respond to local shifts in 'ocean weather' that occur over weeks, hours and minutes, rather than to changes in climate per se, which unfold over years and decades (although long-term climate changes drive the short-term shifts). A handful of studies that attempt to investigate how local physical conditions affect species (including the numbers of individuals and types of species occurring) are beginning to show the value of a more detailed approach[6].

We call on ecologists to rethink their models and experiments. This would enable them to start linking changes in biodiversity to changes in conditions at the scales of space and time that are relevant to individual organisms.

OCEAN WEATHER

To get the most detailed picture of conditions across the ocean's surface and at depth, physical scientists are starting to combine high-resolution *in situ* measurements of temperature, salinity and so on with satellite data. Remote-sensing and continuous monitoring are revealing a highly dynamic environment, even in the open and deep oceans (see 'Watched waters').

For instance, circular currents, or eddies, occur throughout the ocean. Depending on whether they rotate in the same direction as Earth or counter to it, they can provide conditions that are rich or poor in nutrients — different habitats for different phytoplankton and other organisms[7].

The currents arising from eddies extend down 4,000–6,000 metres to the abyssal ocean, as 'benthic storms'. These resuspend sea-floor sediment, creating nutrient-rich regions at depth[8]. Likewise, tides, storms and strong currents affect mixing and change buoyancy throughout the water column, across scales ranging from centimetres to a few metres. This sets the stage for considerable variation in the amount of photosynthesizing life through space and time. And that affects entire food webs.

Nearer to shore, variability is even more dramatic. The temperature can shift by more than 10 °C in one tidal cycle or as the wind displaces surface water and cold water wells up from below (upwelling). Oxygen levels can swing from 0% to 100%, and pH can shift by more than one unit as microbes use up oxygen and as phytoplankton and plants generate it. Microsensors placed near organisms such as mussels have revealed that oxygen, pH and carbon levels can be highly variable, even on small scales of less than 1 millimetre. Extremes of these variables far exceed the projections made by the Intergovernmental Panel on Climate Change under various scenarios for a warming planet[9].

STORM FORCE

Why do ecologists generally ignore such ocean weather as a force that shapes biodiversity?

In our experience, there is a misperception among researchers that highly localized, rapid changes are irrelevant to understanding or predicting biological changes in marine systems. Spikes in temperature and other variables over hours or minutes are often dismissed as being 'extreme' or 'noise'[10]. A greater barrier, however, is obtaining the relevant data in a format that is accessible to biologists.

Satellite measurements of global temperature have been collected by space agencies such as NASA since the 1980s. And today, data on temperature trends, rainfall, cloud cover and other climate phenomena can be downloaded easily. They are available at gridded scales of tens to hundreds of kilometres, and often at yearly or monthly resolutions (see, for example, https://data.nasa.gov). Averaged ocean-relevant data tailored for ecological questions are also available from initiatives such as Bio-ORACLE, run by a team of marine researchers in Belgium, Portugal and Australia. However, the data generated by high-resolution ocean monitoring are much harder to access.

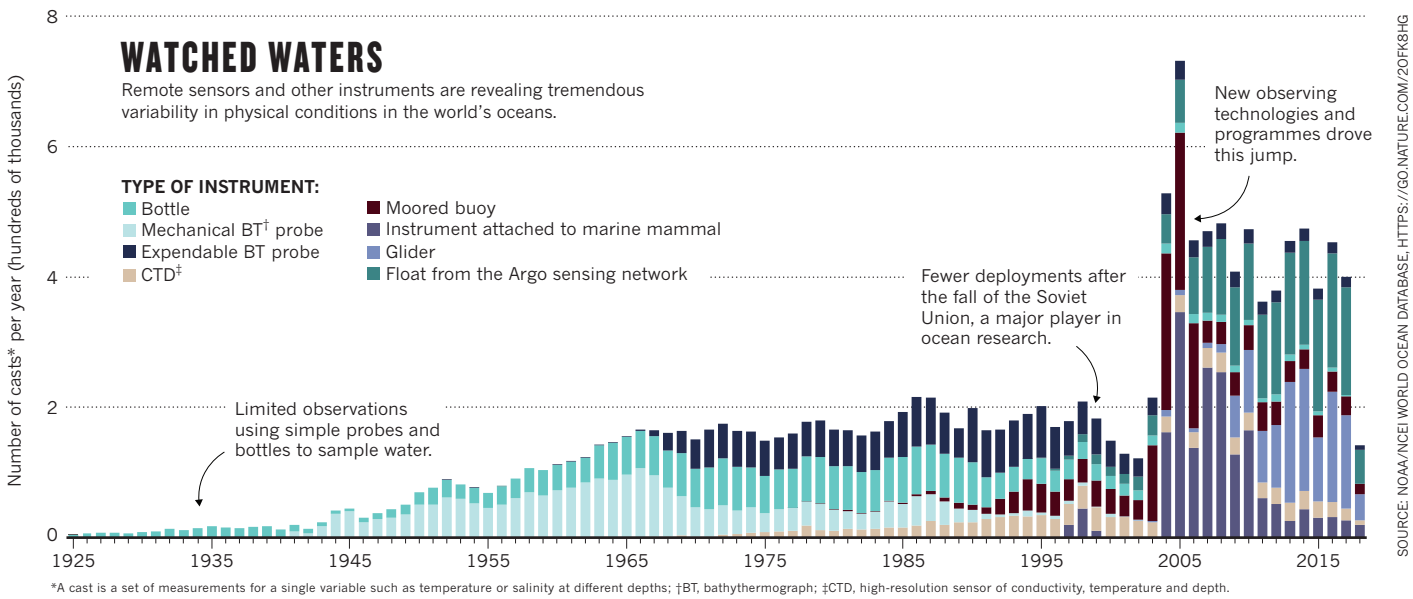
Such monitoring tends to be geographically limited, with the most intensive surveys occurring in waters where nations have economic interests and access. Even when the data have been collected, many ecologists do not have the computational skills or infrastructure to store and manipulate them. When one of us (A.E.B.) recently requested data from a national oceanography institute, for instance, an oceanographer provided a link to many folders. Each folder contained hundreds of files of temperature and other data collected from different time periods — too vast a resource to download on a standard computer.

TURN THE TIDE

This neglect of ocean weather in theory, experimental design and modelling is hampering progress in at least three ways.

Predictions are wrong. When ecologists try to forecast change by running experiments or using macroscale, simulation-based models, physical parameters are generally represented by averages. Such efforts can generate either overly catastrophic projections or excessively optimistic ones.

Ecologists generally agree, for example, that marine species in the tropics and poles will be more vulnerable to the effects of a rise in temperature of 2 °C. Tropical species are already living in the warmest habitats on the planet[11], whereas those at the poles have nowhere else to go[12]. But oceans are not



*A cast is a set of measurements for a single variable such as temperature or salinity at different depths; †BT, bathythermograph; ‡CTD, high-resolution sensor of conductivity, temperature and depth.

DAN COSTA/UCSC

A Weddell seal equipped with a sensor for measuring ocean conductivity, temperature and depth.

warming evenly across the tropics and poles. Some areas are even cooling[13].

Heterogeneity is overlooked. Overall, Earth is losing species. Yet there are huge differences in the rate of loss at the local scale; biodiversity is even increasing in some places[14]. Certain species, populations and individuals can adapt and adjust, and this will lead to surprises. In 2017, hundreds of surveys on the Great Barrier Reef in Australia before and after a mass bleaching event revealed huge variability in how fish species responded to the extreme heat. Some trophic groups, such as herbivores that scrape algae, became less common on the warmest reefs. For others, such as those that feed on plankton, warmer temperatures seemed to benefit populations[15].

Opportunities are missed. Ignoring the variability in ocean systems could limit conservation and management strategies[6,16]. For instance, the concept of climate refugia, where species can shelter from the effects of climate change, has been considered for cooler terrestrial landscapes such as mountain valleys and rivers. Yet marine spatial planning tends to overlook the possibility of refuge sites arising, say, from the upwelling of cooler waters from depth, or from the shade provided by a coral reef. This is largely because ecologists lack the fine-scale data to establish where potential refugia exist.

THE NEXT WAVE

Each stride forward in the physical sciences should translate to improvements in ecologists' predictions of biodiversity change. Major advances in how atmospheric and climate scientists understand ocean processes are rapidly unfolding as a result of improvements in ocean-monitoring technologies, as well as in climate models[17].

Making equivalent progress in the life sciences — in tandem — will require at least three changes.

Acceptance. Ecologists must embrace the fact that the oceans are variable, and consider more carefully the limitations and biases inherent to physical data. Ocean surface temperatures measured by satellites, for example, shed little light on conditions for organisms that live at depth.

In practical terms, this means incorporating variability into ecological models and experiments. This is starting to happen for terrestrial ecosystems. In 2016, for instance, researchers revealed that daily fluctuations in temperature are just as powerful a predictor of changes in the geographical range of frogs, lizards and other organisms as seasonal variation[18].

High-performance computing. Ecologists urgently need ways to access and analyse high-resolution data on environmental variability. They are used to dealing with megabytes of data, but they need to be able to handle terabytes.

Currently, there are various options for accessing high-performance computing. Researchers can apply for cloud-computing grants offered by Microsoft and Google. And some countries, such as Canada, offer cloud resources and training to enable academic institutions to embrace big-data research. The provision of this type of infrastructure and support should be prioritized more broadly.

Crosstalk and collaboration. Much more dialogue is needed between ecologists, physiologists and climate and ocean scientists to aid understanding of what data are required, and in what formats. For instance, sDiv — the Synthesis Centre of iDiv, the German Centre for Integrative Biodiversity Research in Leipzig — runs workshops to foster cross-talk between researchers and kick-start new approaches. This and hundreds of similar efforts could help to bring the relevant researchers together. Dedicated funding for working groups, and for interdisciplinary science in general, will be key.

Only through global collaboration will ecologists be able to obtain a global perspective on ocean weather. There are already some good models for this. The Global Ocean Acidification Observing Network (GOA-ON), for instance, is an international effort to provide highly resolved biogeochemical data on the scale of metres, to enable researchers to optimize models of ocean acidification.

We predict that when biologists engage with the physical and biogeochemical data now becoming available — at scales matched to those of organisms' lives — major shifts will occur in how we conceptualize and manage biodiversity change in the ocean. ■

Amanda E. Bates *is associate professor of ocean sciences at the Memorial University of Newfoundland, St. John's, Canada, and a Canada Research Chair in marine physiological ecology.* **Brian Helmuth, Michael T. Burrows, Murray I. Duncan, Joaquim Garrabou, Tamar Guy-Haim, Fernando Lima, Ana M. Queiros, Rui Seabra, Robert Marsh, Jonathan Belmaker, Nathaniel Bensoussan, Yunwei Dong, Antonios D. Mazaris, Dan Smale, Martin Wahl, Gil Rilov.**
e-mail: abates@mun.ca

1. Helmuth, B. *et al. Sci. Data* **3**, 160087 (2016).
2. Hofmann, G. E. *et al. PLoS ONE* **6**, e28983 (2011).
3. Naveira Garabato, A. C. *Phil. Trans. R. Soc. A* **370**, 5480–5511 (2012).
4. Tizard, T. H. & Moseley, H. N. *Narrative of the Cruise of HMS* Challenger*: With a General Account of the Scientific Results of the Expedition* (Vol. 2). (HM Stationery Office, 1882).
5. Petsko, G. A. *Genome Biol.* **12,** 112 (2011).
6. Helmuth, B. *et al. Clim. Change Responses* **1**, 10–20 (2014).
7. Condie, S. & Condie, R. *Glob. Ecol. Biogeogr.* **25**, 1264–1277 (2016).
8. Gardner, W. D., Richardson, M. J. & Mishonov, A. V. *Earth Planet. Sci. Lett.* **482**, 126–134 (2018).
9. Hurd, C. L. *et al. Glob. Change Biol.* **17**, 3254–3262 (2011).
10. Dowd, W. W., King, F. A. & Denny, M. W. *J. Exp. Biol.* **218**, 1956–1967 (2015).
11. Comte, L. & Olden, J. D. *Nature Clim. Change* **7**, 218–722 (2017).
12. Flynn, E. E., Bjelde, B. E., Miller, N. A. & Todgham, A. E. *Conserv. Physiol.* **3**, cov033 (2015).
13. Josey, S. A. *et al. Ann. Rev. Mar. Sci.* **10**, 475–501 (2018).
14. Dornelas, M. *et al. Science* **344**, 296–299 (2014).
15. Stuart-Smith, R. D., Brown, C. J., Ceccarelli, D. M., & Edgar, G. J. *Nature* **560**, 92–96 (2018).
16. van Hooidonk, R., Maynard, J. A. & Planes, S. *Nature Clim. Change* **3**, 508–511 (2013).
17. Hewitt, H. T. *et al. Geosci. Model Dev.* **9**, 3655–3670 (2016).
18. Chan, W-P. *et al. Science* **25**, 1437–1439 (2016).

A full list of author affiliations accompanies this Comment online (see go.nature.com/2m8acbj).

Musk oxen on Russia's Wrangel Island.

SERGEY GORSHKOV/NPL

CONSERVATION

Life in the deep Arctic wild

Huw Lewis-Jones revels in a study of field biology at the frontiers of climate change.

I write this from a Russian icebreaker on a voyage to the North Pole. On board are passengers from many countries, their reasons for heading into this most extreme of environments ranging from the deeply personal to the mildly boastful. One wants to "see the ice before it's too late" (this winter, the maximum extent of Arctic sea ice was the second-lowest on record). For most, wildlife is a major draw, and king of all is the polar bear. *Ursus maritimus*, the 'sea bear', has become the symbol of the high Arctic and the poster-child for ecological crisis.

Many more species of the cold wilderness share their uncertain future, reveals conservation biologist Joel Berger in *Extreme Conservation*. Berger has trailed yaks (*Bos mutus*) on the Tibetan Plateau and the proboscis-swinging saiga (*Saiga tatarica*) through Mongolia's Gobi Desert. He has tracked one species for more than a decade. "If polar bears are the face of climate change," he writes, "muskoxen are the heart." This hairy, horned Arctic bovid (*Ovibos moschatus*), neither an ox nor a maker of musk, once roamed with woolly mammoths. Yet we still know very little about its physiology, reproduction, predation or food sources; least of all about its adaptive capacity.

Extreme Conservation: Life at the Edges of the World
JOEL BERGER
University of Chicago Press (2018)

UP CLOSE AND PERSONAL

To study the musk-ox life cycle demands considerable effort and expense. The hostile conditions are especially complex in winter, with severe blizzards and temperatures far below freezing. But by helicopter

and snowmobile, the indefatigable Berger perseveres. He mostly eschews radio collars for less intrusive data-collection methods, such as analysing frozen dung for stress hormones, and photogrammetry — assessing an animal's health from photographs.

Musk oxen were extirpated from Alaska during the quarter of a century after the United States purchased the land from Russia in 1867. In the 1930s, efforts began to re-introduce the species to its former range: juveniles captured in Greenland were transported by ship, rail and boat by way of Norway to Nunivak Island in the Bering Sea. Descendants of the 27 animals that completed the trip were then moved throughout Alaska, and even to Russia's Wrangel Island, north of Siberia; the United States offered these animals to the then-Soviet Union during the cold war as a token of friendship — a kind of panda diplomacy. Efforts to maintain and restore populations of the enigmatic creatures are now thwarted by the effects of global warming. Berger has made it his life's mission to understand the mechanisms through which these changes affect populations.

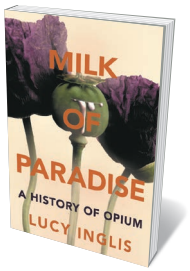
"If polar bears are the face of climate change, muskoxen are the heart."

He is an excellent guide, a respected ecologist and a gifted storyteller. It's an important combination in environmental advocacy: scientists who can tell compelling stories can elevate their research outputs. His previous books include *Wild Horses of the Great Basin* (1986), the *Horn of Darkness* (1997; written with Carol Cunningham) and the 2008 *The Better to Eat You With*. In his 2008 book, he tracks cultures of fear in animals across continents and climates, from elk and wolves in the US Yellowstone National Park to moose coexisting with tigers and bears in Asia. He opens our eyes to what it takes for such animals to cling to the edges of existence.

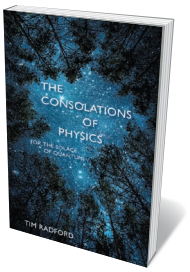
TRAPPED IN THE ICE

Independent studies by agencies such as NASA and the US National Oceanic and Atmospheric Administration have shown, year on year, that global average surface temperature is rising and that warming patterns are most extreme in the polar regions. One of the most troubling recent observations in the Arctic is that the number of winter days when temperatures don't drop below freezing is increasing. Precipitation falls as rain, melting snow on the ground; when the real cold returns, lichens and grasses are encased in steel-hard ice, impenetrable to hooves. In the early 2000s, Berger relates, such an event on Canada's Banks Island led to the deaths of some 20,000 musk oxen. And other extreme ▶

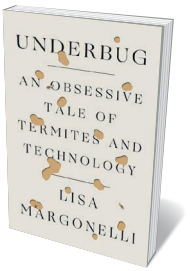
Books in brief



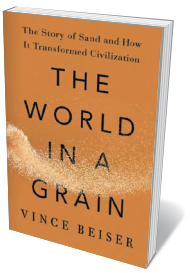
Milk of Paradise
Lucy Inglis MACMILLAN *(2018)*
It has ignited brutal wars, addicted millions and transformed medicine. Opium (the sap of the poppy *Papaver somniferum*), as cultural historian Lucy Inglis's magisterial chronicle reveals, has been used and abused for millennia, wending its way from ancient Mesopotamia to Asia and Victorian Britain as a sedative and recreational drug. The emergence of morphine and heroin in the nineteenth century sparked pharmacological advances, as well as illicit trade, conflicts and the US opioid crisis. As Inglis observes, "we must never forget that this is a battle fought only with ourselves".



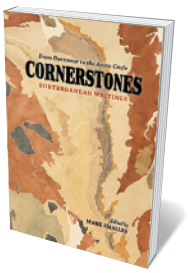
The Consolations of Physics
Tim Radford SCEPTRE *(2018)*
This beautifully crafted "love letter to physics" by science writer Tim Radford hinges on the Voyager mission, launched in 1977 to study the outer Solar System and now heading out to the heliopause and beyond. Radford (a frequent contributor to these pages) finds solace in Voyager as a grand cooperative effort in a world too often at war. His deft narrative interweaves discoveries such as the Higgs boson, the Hubble Deep Field and gravitational waves with Dante Alighieri's epic fourteenth-century poem *The Divine Comedy*, which intuited the laws of motion found by Galileo Galilei some 300 years later.



Underbug
Lisa Margonelli FARRAR, STRAUS AND GIROUX *(2018)*
Termites are not just the destructive force that homeowners know and hate — "architects of negative space", as environmental writer Lisa Margonelli wittily puts it. They also comprise a kind of entomological three-ring circus, and this round-up of research on the eusocial insects is a ticket to the show. Their guts teem with wood-degrading enzymes that could revolutionize biofuels; the convoluted interiors of their mounds reveal astonishing group behaviours; they even engineer ecosystems by revitalizing soils. This is a wild ride through a hidden microcosmos stretching from Australia to Namibia.



The World in a Grain
Vince Beiser RIVERHEAD *(2018)*
Our world is built on sand: it's in everything from silicon chips to concrete. Currently, the global construction industry consumes some US$130 billion worth of the stuff annually — so much so that this ubiquitous natural resource is running out. Journalist Vince Beiser's eye-opening study clarifies the science and the huge role of sand in heavy and high-tech industry. Perhaps most compelling is his exposé of sand mining, which obliterates islands, destroys coral reefs and marine biodiversity, and threatens livelihoods. A powerful lens on an under-reported environmental crisis.



Cornerstones
Mark Smalley (editor) LITTLE TOLLER *(2018)*
This stirring collection of 22 essays on Earth's bedrock intermingles geology and our cultural responses to stone, from flint tools and megaliths to Gothic cathedrals. Editor Mark Smalley has assembled a group of stellar contributors. Novelist Sarah Moss muses over igneous whinstone under Hadrian's Wall in Northumbria, UK; environmental writer Jason Mark ponders ethereal Arctic uplands at risk from oil drilling; poet Fiona Hamilton sees bricks as compressed time and energy. Rock, Smalley reminds us, is both thing and metaphor, helping us feel our way "towards the intangible". **Barbara Kiser**

events are becoming more common. In 2011, more than 50 were entombed in ice during a Chukchi Sea winter tsunami (*ivuniq* in the Iñupiat language of northwestern Alaska); elsewhere, rapid freezing has entrapped whales and sea otters.

Freak weather and storms of great intensity are aspects of an alarming reality not confined to the Arctic. Worldwide, biodiversity loss is accelerating at an unprecedented rate. For Berger and his peers, it has become a moral obligation to apply their knowledge to ecological well-being. "Doing science is not conservation," Berger writes. "Donning a human face, inspiring people to care, engaging people who listen, and ultimately persuading decision makers to act is."

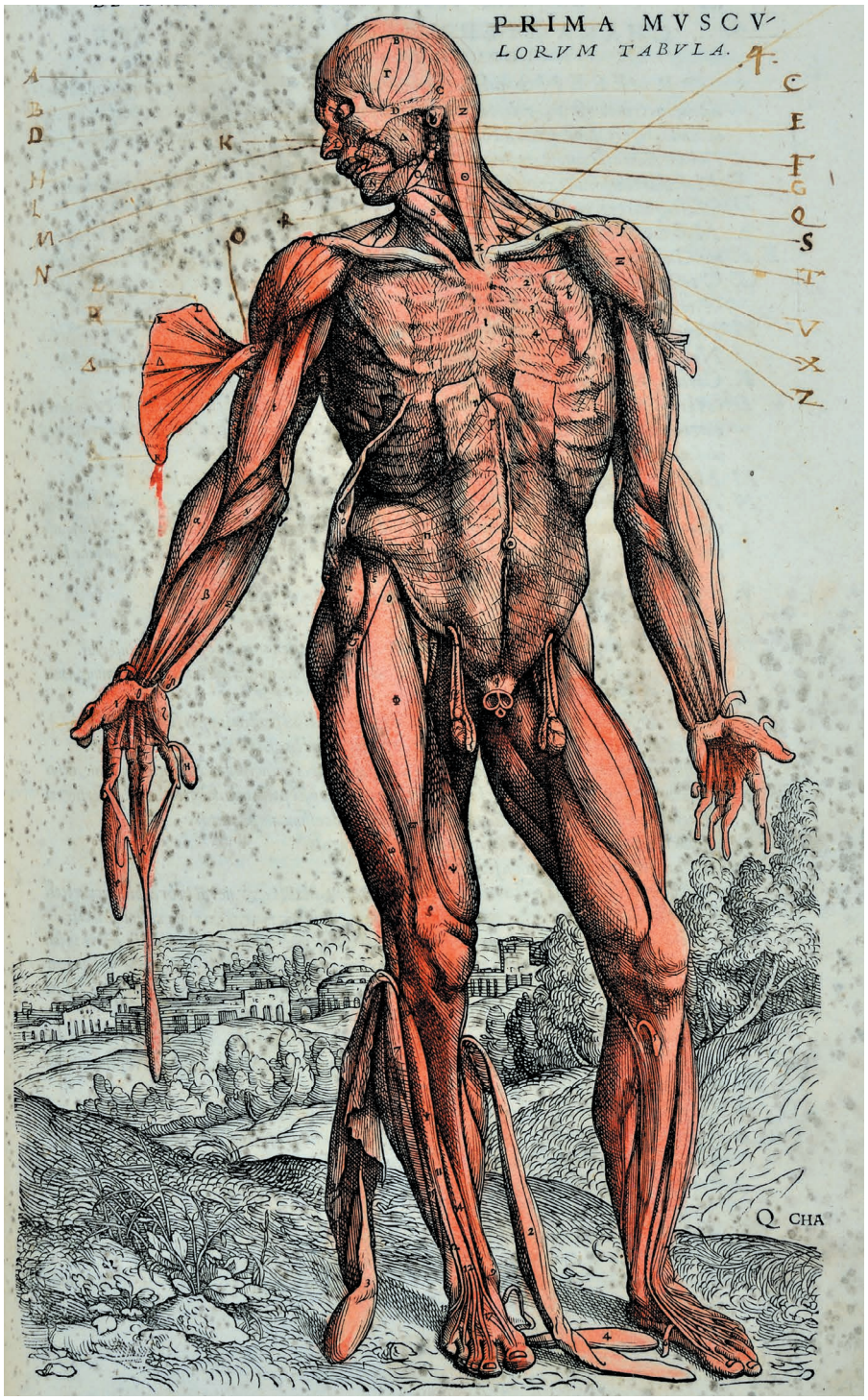
Berger's methods can be eccentric. He pitches carnivore dung at moose, baseball-style, to see whether they respond to the scent, or wears a polar-bear costume made from cloth and styrofoam to get close to musk oxen and study their reactions. "He is the hairy-arsed action-man academic," *The Times* newspaper once wrote. His data, gleaned primarily from northern latitudes and extreme heights in the Himalayas, are in my view all the more insightful for that.

"The more adapted a species has become to its ecological niche, the more devastating climate change can be for it."

What Berger's fieldwork shows us is that the more adapted a species has become to its ecological niche, the more devastating climate change can be for it. With receding sea ice, polar bears find it more difficult to hunt seals, their favoured prey, and now forage more widely onshore, for example on eggs of migratory waterfowl. The melt shrinks their hunting season and the time they have to rest and breed, which they normally do on sea ice. "Life at the extremes is more challenging than ever," he writes, "and the need for action, for solutions, has never been greater."

We need to remember, too, that cold-adapted species have survived "across thousands of generations", Berger notes. A fraction of that time is left for us to curb the impacts of climate change. ■

Huw Lewis-Jones *is an environmental historian and expedition guide, working regularly across the Arctic and Antarctica.*
Twitter: @polarworld



BIBLIOTECHA CIVICA ROMOLO SPEZIOLI, FERMO. CALL NO.: 1E7N.1582

A 'muscle man' from *De humani corporis fabrica*, annotated by a reader.

MEDICAL HISTORY

Sex, religion and a singular anatomist

Andreas Vesalius's images baffled many early on, reveal **Dániel Margócsy**, **Mark Somos** and **Stephen N. Joffe**.

The Renaissance anatomist Andreas Vesalius's *De humani corporis fabrica* ('On the fabric of the human body') is a foundational work of medicine in the West. Its more than 200 woodcuts revolutionized how people pictured the human body, flayed and cut to reveal musculature, nerves, organs and bones. Even now, 475 years after it was first published, the bold images of skeletons and skinless 'muscle men' in sinuous poses (by illustrator Jan Steven van Calcar) beguile.

More than 700 copies survive from the 1543 and 1555 editions, which Vesalius supervised. Of these, roughly two-thirds contain comments in the margins, bizarre doodles, and coloured-in and even defaced images, as we reveal in our book *The* Fabrica *of Andreas Vesalius*. Early readers, on evidence, studied Vesalius's treatise diligently, yet had no compunction about scribbling in a hugely expensive volume.

Looking deeper, the marginalia tell two stories. One is that some found the images baffling, and attempted to clarify them in innovative ways. Another is that the pious found the figures' necessary nudity scandalous, and felt impelled to weigh in with ink and scissors. Our study of the reactions of hundreds of readers has taught us that medical communities do not always adopt innovative solutions quickly, even when they are presented in such an elegant format as the *Fabrica*. It takes time to get used to novelty. And we have learnt that even the most ingenious scientific minds can fail to predict how political and religious institutions will respond to their work.

The *Fabrica*'s early readers were the first generation of physicians and surgeons in Europe to face the daunting task of using detailed printed images to identify the organs of the body and learn about human physiology. Vesalius and van Calcar faced challenges of their own. The *Fabrica*'s image of the branching portal vein, which carries blood from the intestines to the liver, is highly complex — and does not quite succeed. It is almost impossible, for example, to single out the haemorrhoidal vein. (At the time, this was important because the vein was supposed to be the cause of both menstruation and haemorrhoids, thought to be analogous processes that purged corrupted blood from the body.) Thus, in a copy now in the library of Queen's College at the University of Oxford, UK, someone used a quill and red ink to colour in this meandering vein, like a child playing a maze game.

In a copy once owned by Nuremberg physician Georg Palma, an intricate image of the brain is 'enhanced'. Palma painted six pairs of cranial nerves different hues in watercolour and used the same colours to underline the corresponding pairs in the text on the following page.

Even Vesalius realized that his images could be confusing, and devised an ingenious method to explain them. A letter or number was printed onto the image of each body part, with a separate key. Unfortunately, the characters were often too small to pick out against the swirling background. Some frustrated readers underlined, highlighted, enlarged or repeated the characters in the margins. On one muscle man, for instance, the tiny character identifying a thigh muscle was barely visible, and a confused reader queried desperately whether it was the Greek letter μ or the Roman letter u.

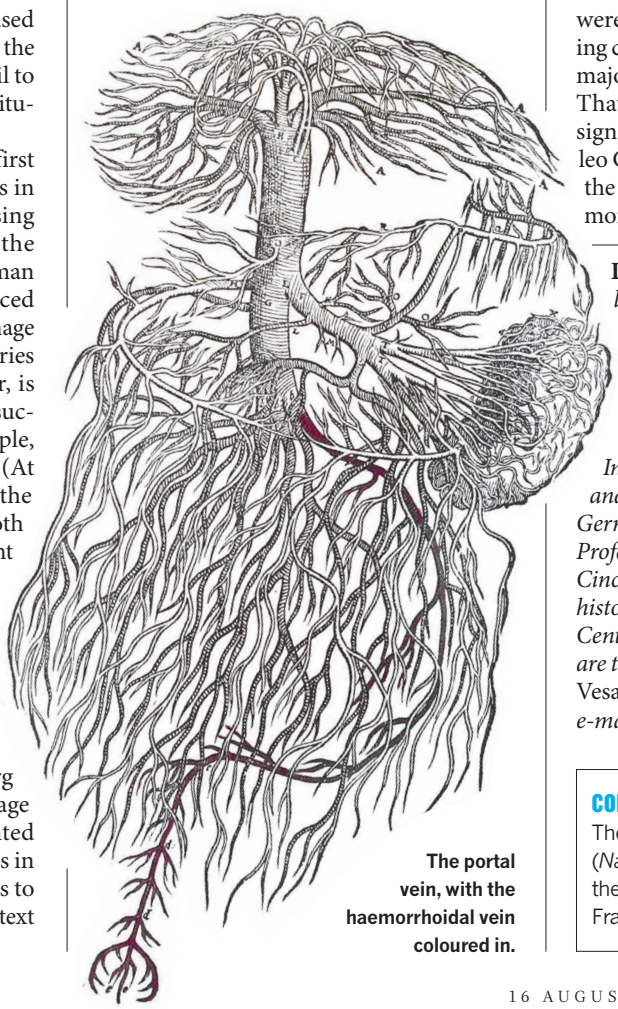
"Even the most ingenious scientific minds can fail to predict how political and religious institutions will respond to their work."

Faced by such challenges, many medics might have given up on the images. Indeed, when we reconstructed what early modern readers and scholars found fascinating about the *Fabrica*, it was evidently the text. The clear majority of sixteenth- and seventeenth-century readers who annotated the book focused on that and left no traces of having engaged with the illustrations. Sixteenth-century reviews of the *Fabrica* confirm this impression, because they tended to discuss only the text.

This is no surprise. The *Fabrica*'s scholarly readership was trained in the traditions of Renaissance humanism, which put a strong emphasis on textual analysis. Even if they found it difficult to interpret visual information, medical practitioners were expert at making sense of long Latin texts. Furthermore, the body's 'interior universe' had hardly been mapped. Even today, it is difficult to make sense of images of internal organs if you've never seen a dissected body, and radiologists need years of training to interpret X-rays or magnetic resonance imaging scans.

If images were not that helpful for understanding the body, what was their purpose? For Church authorities in the period, the answer was clear. They argued that such figures held an erotic appeal because they showed the genitals — and so should be censored. The first version of the *Index librorum prohibitorum*, the list of books banned by the Catholic Church, came out in 1559, and it did not mince words about 'licentious' books. That included tomes on anatomy. Many owners of the *Fabrica* felt that they had to paint aprons on the muscle men (as in the copy once owned by the Jesuit College of Bourges in France), or snip the offending parts out.

Only a minority of copies of the *Fabrica* were so treated. We checked every surviving copy, and found the images intact in the majority owned by Catholics in the period. That invasive censorship happened at all signals that, at least until the trials of Galileo Galilei in the early seventeenth century, the Church found anatomical illustrations more dangerous than heliocentrism. ■



THE PROVOST AND FELLOWS OF THE QUEEN'S COLLEGE, OXFORD

The portal vein, with the haemorrhoidal vein coloured in.

Dániel Margócsy *is university lecturer in the Department of History and Philosophy of Science at the University of Cambridge, UK.* **Mark Somos** *is Alexander von Humboldt Foundation Fellow and senior research affiliate at the Max Planck Institute for Comparative Public Law and International Law in Heidelberg, Germany.* **Stephen N. Joffe** *is Esteemed Professor of Surgery at the University of Cincinnati in Ohio and visiting professor in history of medicine at Cedars-Sinai Medical Center in Los Angeles, California. They are the authors of* The *Fabrica* of Andreas Vesalius *(2018).*
e-mail: margocsy@gmail.com

CORRECTION

The Books & Arts article 'Summer books' (*Nature* **559**, 328–330; 2018) misnamed the author of *On Bullshit*; he is Harry Frankfurt.

Correspondence

Shock prize announcement

This year's Gruber Cosmology Prize — the most prestigious in the field — went to the European Space Agency's Planck satellite observatory team for its precise measurement of the Universe's contents and contours. The US$500,000 prize will be awarded on 20 August in Vienna. The Planck team has more than 300 members, of whom about one-fifth are women. Yet the collaboration has indicated that the team's half-share of the prize money (two principal investigators share the other half) would be divided between 43 senior members of the collaboration, all of whom are men. Although the number of recipients has to be limited and the prize money might end up being pooled, it is remarkable that this situation has arisen in 2018.

That all Planck's female scientists have even temporarily been deemed unworthy of controlling a share of the prize is unwelcome news, especially to the many of us trying to tackle the under-representation of women in astronomy.

Olivier Berné *Research Institute in Astrophysics and Planetology, University of Toulouse, France.*
olivier.berne@irap.omp.eu

Speed up global ban on trans fats in foods

I suggest that the food industry should be subject to a time limit for removing hazards identified in the global food system (see L. Haddad *Nature* **556**, 19–22; 2018). For example, we have known for decades that industrially processed *trans*-fatty acids (TFAs) in food are a risk factor for cardiovascular disease. Although TFAs can be removed from the food supply efficiently, in many countries these still persist (see, for example, S. Stender *et al. BMJ Open* **6**, e010673; 2016).

Denmark has been leading the fight against TFAs since 2004. And seven years have been lost since the European Union issued its food-labelling regulation in 2011, which would have been an opportunity to tackle TFAs. Although a TFA ban is still on the EU agenda (see go.nature.com/2okkffs), taking action is up to individual states — for example, TFAs are no longer permitted in Slovenia.

By contrast, a ban on partially hydrogenated vegetable oils, the source of TFAs in food, has just come into effect in the United States; Canada will follow next month. And the World Health Organization this year made elimination of TFAs by 2023 the highest priority in its 'REPLACE' action programme (see go.nature.com/2vintqf).

Igor Pravst *Nutrition Institute, Ljubljana, Slovenia.*
igor.pravst@nutris.org
Competing interests declared; see go.nature.com/2nyeabf

Undergrad research: begin at the start

I question the common practice of training undergraduate students in research only during their final year at university (see J. Ankrum *Nature* http://doi.org/gdwps2; 2018). For an honours thesis, a student in Canada typically spends 4 months learning technique to inform 4 months — or around 350 hours — of effective research. Instead, my lab trains undergraduates in research methods throughout their degree, so that they have about 1,200–1,600 hours of hands-on research experience by the time they graduate.

We use a vertical peer-mentoring system in my synthetic-chemistry lab: graduates and postdocs oversee undergraduate projects, assisted by experienced undergraduate researchers, who help to train the new undergraduates. The lab currently hosts 25 undergraduates, with an annual intake of around 7 promising first-year students. They spend their first year learning and developing techniques, their second testing and troubleshooting chemical reactions, their third designing and implementing small research projects, and their final year producing their thesis.

John Trant *University of Windsor, Canada.*
j.trant@uwindsor.ca

Tackling AI impact on drug patenting

Initiatives are already under way to avoid ill-considered moves concerning the impact of artificial intelligence (AI) on drug patenting (see L. Heuer *Nature* **558**, 519; 2018).

Heuer mentions some of the issues. For example, he foresees problems over whether to designate the algorithm or its programmer as the inventor, and whether a drug discovered through machine-learning methods would be patentable.

In the United States, at least, some of these issues are currently clear. For example, US patent law states that "a person shall be entitled to a patent", and an algorithm is not a person. It also states that "patentability shall not be negated by the manner in which the invention was made". More generally, it is insufficient to assert that just because an AI could arrive at a particular solution, then that solution must be obvious.

However, a serious problem for pharmaceutical companies is that, according to US law, only people can make the inventive step in patents. In practice, it is likely that algorithms are making many of those steps, raising questions about the validity of these patents in the United States. We welcome efforts to arrive at a consensus over such dilemmas by the robotics research community (see go.nature.com/2onhgcb), intellectual-property professionals (see go.nature.com/2oiwh4c), the European Commission and the European Patent Office.

Ross D. King *University of Manchester, UK.*
Patrick Courtney *tec-connection, Konstanz, Germany.*
ross.king@manchester.ac.uk

Junior reviewers jump into the pool

As members of the Association of Polar Early Career Scientists (APECS), we participated in a group review of the upcoming report on the ocean and cryosphere from the Intergovernmental Panel on Climate Change (IPCC). Our analysis compared well with reviews by more senior scientists (see also L. van der Veer *et al. Clim. Change* **125**, 137–148; 2014). Early-career scientists are an untapped source of peer reviewers who, in our view, could be deployed as successfully on journal manuscripts as on large reports.

We encourage other early-career scientists to engage in individual and group reviews, such as those organized by APECS, including the second review of the IPCC ocean and cryosphere report taking place later this year.

Expanding the reviewer pool in this way would benefit the scientific community by mitigating the review burden (see, for example, M. Kovanis *et al. PLoS ONE* **11**, e0166387; 2016). There would be career advantages for junior researchers who were accomplished reviewers. And they would gain insight into improving the preparation and presentation of their own papers.

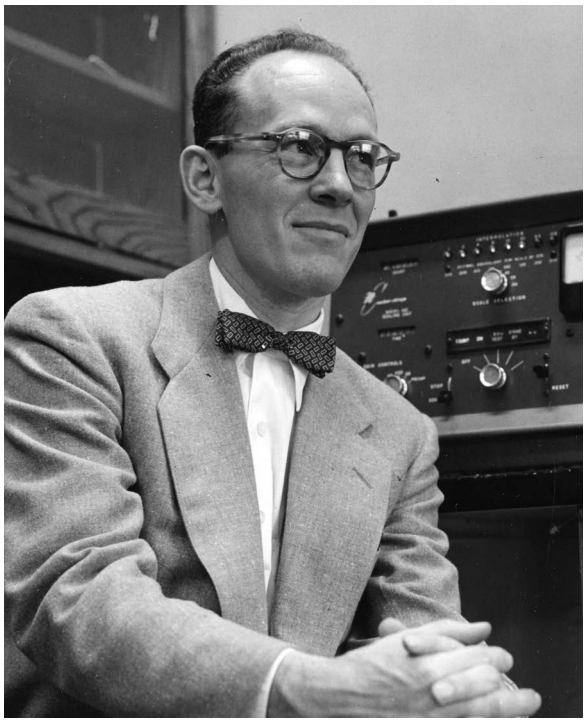
Comprehensive reports such as those compiled by the IPCC provide a means for the scientific community to reach the public. Such engagement is becoming increasingly important, so early-career researchers must learn to contribute to it effectively.

Mathieu Casado* *Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany.*
mathieu.casado@gmail.com
**On behalf of 12 co-signatories (see go.nature.com/2mdtryf for full list).*

Paul Delos Boyer

(1918–2018)

Nobel laureate in chemistry who discovered nature's smallest rotary motor.



UNIV. MINNESOTA ARCHIVES

Paul Boyer was approaching the finish line of his career when he risked everything with a jaw-dropping proposal. He addressed one of the most important, as-then-unanswered questions in biochemistry: how do cells use an electrochemical gradient to form the molecule adenosine triphosphate (ATP), which drives the energy-requiring processes that are essential to life?

More than 90% of the ATP formed in our cells is catalysed by an enzyme called ATP synthase. Boyer's remarkable proposal for how the enzyme works was that ATP synthase functions as a tiny molecular motor. Just as electric motors spin when electrons flow through a potential gradient, a flow of protons (hydrogen ions) across an electrochemical gradient generated by respiration causes the core of the ATP synthase to spin relative to its surrounding catalytic subunits. Because the core is asymmetric, its rotation causes the surrounding subunits to change their shape, thus disrupting the tight binding site where ATP is formed, allowing its release.

Boyer died on 2 June, eight weeks shy of his 100th birthday. Born in Provo, Utah, he studied chemistry at Brigham Young University, also in Provo. After marrying his college sweetheart, Lyda Whicker, he headed in 1939 to the biochemistry department at the University of Wisconsin–Madison for his PhD degree. After postdoctoral training at Stanford University, California, in 1946 he accepted a faculty position at the University of Minnesota, St Paul. In 1963 he returned to California to head the biochemistry division of the chemistry department at the University of California, Los Angeles (UCLA), where he spent the rest of his career.

Boyer's revelations about ATP synthase occurred in three steps and involved insights that defied dogma. I was fortunate to be a postdoc in Paul's lab when he came up with the first of these.

We were attending a UCLA seminar in 1972 when I noticed that he wasn't paying attention to the speaker. Afterwards, Paul approached us in a very excited state. This was surprising because he was known for his calm demeanour. He confessed that he had spent the hour thinking about old unexplained data. He asked: "What would you say if I told you that it doesn't take energy to make ATP at the catalytic site of ATP synthase," (as was universally held at the time) "but rather that it takes energy to get ATP off the catalytic site?" This was a eureka moment.

In other words, tightly bound ATP forms spontaneously at catalytic sites on the synthase. Energy from the electrochemical gradient is used to alter the surrounding structure, which disrupts the tight site, allowing ATP release.

As is often the case with transformational ideas, early reactions were negative. When the *Journal of Biological Chemistry* rejected our manuscript containing data supporting this concept, Boyer told me without animosity that he could see why they would do that — "It was a very striking claim." The work was published in 1973 in the *Proceedings of the National Academies of Science*.

A second feature of the mechanism, on which he published a 1977 paper, involved recognizing that the multiple catalytic sites of the synthase coordinated with one another: the finished ATP product is released only when ATP precursors bind at another, adjacent site.

Finally, in 1981–82, Boyer proposed that rotation of the asymmetric core of the synthase, driven by an electrochemical gradient, led to the necessary changes in the surrounding catalytic subunits.

Subsequently, all three concepts received strong experimental support from numerous labs. This included results of X-ray crystallographic studies conducted by John Walker at the UK Medical Research Council's Laboratory of Molecular Biology in Cambridge; he was also a recipient of the 1997 Nobel Prize in Chemistry.

Boyer had a gift for extracting insights from data. But he also drew on knowledge acquired over years of reading beyond his field. As editor of 19 volumes in a series of books called *The Enzymes*, and as co-editor and editor of the *Annual Review of Biochemistry*, he followed all the major advances in enzymology over several decades. He also gained an edge by coming to biochemistry as a chemist, not as a biologist as was common in the 1940s. His firm understanding of kinetics and thermodynamics played an important part in his discoveries.

Boyer was a highly effective trainer of young scientists. Mentoring by example, he exhibited an incredible work ethic, a love for science and a pure joy in discovery. He referred to his favourite enzymes as friends. He set high expectations while providing a supportive and stimulating environment. Even after you left his lab, he looked for opportunities to help you advance your career.

Paul was also a role model for how to live your life with integrity, respect for others and kindness. He showed exceptional civility in a field known to be contentious (shouting matches were common at scientific meetings). In all the years we worked together, I never heard him raise his voice in anger or offer a personal criticism of any of his competitors. When faced with a difficult situation, I still ask myself: how would Paul Boyer handle this problem? The answer is always: with grace and generosity. ■

Richard L. Cross *is SUNY distinguished professor emeritus of biochemistry and molecular biology at Upstate Medical University in Syracuse, New York, USA.*
e-mail: crossr@upstate.edu

ORGANIC CHEMISTRY

A new twist on catalytic teamwork

Catalysts working in pairs can promote more-effective reactions than can the same catalysts used sequentially. The coupling of an enzyme with a light-activated catalyst offers great potential for organic synthesis. SEE LETTER P.355

NICHOLAS J. TURNER

The development of catalytic reactions is a dominant theme in chemistry, especially in industry, where major efforts are under way to develop large-scale chemical processes that are sustainable and avoid producing unnecessary waste[1]. Chemical reactions can be accelerated using many types of catalyst, including metals (or their salts or complexes), small organic molecules, enzymes and light-activated catalysts. Catalysts of all types have advanced to the extent that two or more catalysts can be combined to promote cascade reactions — interconnected transformations, carried out in a single operation, to yield products with selectivities that would be difficult to achieve using the catalysts independently in sequential steps. On page 355, Litman *et al.*[2] report that the combination of an enzyme with a light-activated catalyst starts a cascade reaction that produces compounds that are versatile intermediates for organic synthesis.

The use of combinations of catalysts could potentially lead to step changes in the efficiency of chemical processes[3]. Certain combinations of enzymes with small-molecule organic catalysts or transition-metal catalysts have been of particular interest — in part because the chemistry mediated by these different catalyst types is highly complementary, and also because water is used as the main solvent, thus avoiding environmentally harmful organic alternatives. Moreover, such combinations can open up synthetic routes for constructing molecules that would not otherwise be possible[4]. The emergence of light-activated catalysts (photocatalysts) in the past few years has presented opportunities for the development of systems that combine enzymes with photocatalysts.

Enter Litman *et al.*, who have used just such a combination to promote reactions of alkenes — organic compounds that contain carbon–carbon double bonds. Many alkenes can form as isomers, known as (*E*)- and (*Z*)-isomers, which differ in the geometrical arrangement of groups attached to their double bond (Fig. 1a). Although methods exist that allow just one isomer of an alkene to be produced during synthesis, it is often cheaper and easier to prepare alkenes as mixtures of (*E*)- and (*Z*)-isomers. But using such mixtures can be problematic. Alkenes are often chemically reduced during organic syntheses — that is, the carbon–carbon double bonds are converted into single bonds. But the reduction of both types of isomer together typically yields products known as stereoisomers, which have different spatial arrangements of groups attached to a specific carbon atom and can be difficult to separate.

The ideal solution to this problem would be to combine a catalyst that converts (*E*)-alkenes into (*Z*)-alkenes (and vice versa) with a second catalyst that reduces only one of the two alkene isomers, thus yielding only one stereoisomer. For example, if the second catalyst reduces only (*E*)-alkenes, then it initially catalyses the reduction of any of that isomer found in the original alkene mixture, and then goes on to reduce any (*E*)-isomer produced from the (*Z*)-isomer by the first catalyst. Both isomers of the alkene mixture are thus eventually consumed to make the same product. This is exactly what Litman *et al.* report in the current work.

The authors' catalytic system builds on previous work[5] that reported the use of iridium-based photocatalysts to interconvert the (*E*)- and (*Z*)-isomers of a range of different alkenes. In the current study, Litman and colleagues combined analogues of those iridium photocatalysts with ene reductase enzymes, which reduce alkenes and are generally (*E*)-selective, although the authors also tested a (*Z*)-selective ene reductase in their system.

The researchers optimized various

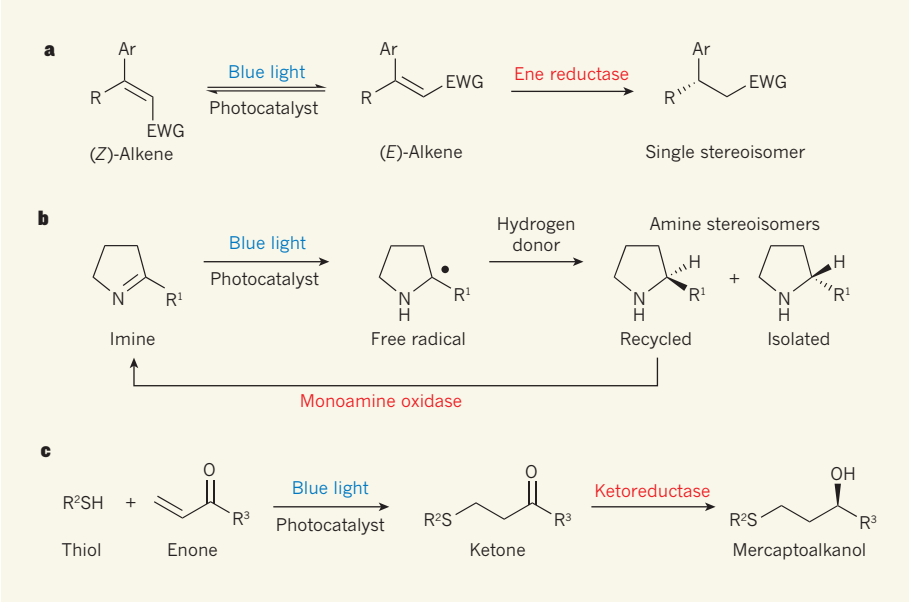


Figure 1 | Combinations of enzymes and light-activated catalysts enable organic reactions. **a**, Litman *et al.*[2] used a light-activated catalyst (a photocatalyst) *in situ* with an enzyme (ene reductase) to catalyse the conversion of mixtures of (*E*)- and (*Z*)-isomers of alkenes into products that form predominantly as a single stereoisomer (an isomer that contains a particular spatial arrangement of bonds). The photocatalyst promotes the interconversion of the isomers, whereas the enzyme reduces only the (*E*)-isomer. The hashed bond extends into the plane of the figure. **b**, A different photocatalyst has been used[6] to convert imines into free radicals (dot indicates an unpaired electron), which accept a hydrogen atom from a donor to form amines as a mixture of stereoisomers. The enzyme monoamine oxidase recycles one stereoisomer back to the imine, which is therefore eventually all converted into the other stereoisomer. The solid-wedge bond extends out of the plane of the figure. **c**, Another photocatalyst promotes[7] the addition of thiols to enones to form ketones, which can be reduced *in situ* by a ketoreductase enzyme to form mercaptoalkanols. R groups represent a variety of organic groups; Ar, aryl group; EWG, electron-withdrawing group.

parameters of their reactions, including the concentrations of the iridium catalyst, the enzyme and the enzyme's cofactor. They developed a system that reduces mixtures of (*E*)- and (*Z*)-isomers of alkenes to form a single stereoisomer, in multi-milligram quantities (Fig. 1a). The authors went on to convert the stereoisomer into a variety of biologically active molecules and key intermediates that have been used to prepare such molecules, thereby highlighting the potential application of their chemistry for preparative organic synthesis.

As Litman *et al.* point out, photocatalytic reactions typically occur at or near to room temperature, which makes them compatible with the thermal requirements of enzymatic systems. Photocatalysts also often work through mechanisms (such as outer-sphere electron transfer and energy transfer) that generate intermediates that are stable in the presence of water and tolerant of the chemical groups found in enzymes. Therefore, photocatalysts in general might be particularly suitable for being combined with enzymes for synthetic reactions.

This compatibility of photo- and enzymatic processes has been exploited in two other studies published earlier this year. In the first[6], a water-soluble iridium catalyst was combined with the enzyme monoamine oxidase (MAO-N) to convert racemic mixtures (one-to-one mixtures of mirror-image stereoisomers known as enantiomers) of amine compounds into a single enantiomer (Fig. 1b). This process begins by generating a highly reactive free radical from a starting material (an imine). The radical is then converted *in situ* to a racemic mixture of amines. MAO-N recycles only one of the enantiomers back into the imine, and the whole process repeats until all of the imine has been converted into the enantiomer that is not the substrate for MAO-N. In the second study[7], a photocatalytic reaction of thiols with enones was used to generate ketone intermediates that were reduced *in situ* with a ketoreductase enzyme, yielding products known as mercaptoalkanols enantioselectively (Fig. 1c).

As with all enzymatic systems, the reaction scope and scalability of Litman and colleagues' transformation will determine the extent to which it finds practical applications. For example, the alkene substrates reported in the paper are linear molecules that bear aryl groups (structural units that contain benzene rings; a substrate bearing an aryl group known as a pyridine ring is also reported). It will be interesting to see whether the chemistry can be extended to cyclic and non-aryl-bearing substrates. Moreover, the concentration of substrates used in the reactions is currently lower than would be needed for industrial processes. It remains to be seen whether the photocatalyst and enzyme will work at industrially useful substrate concentrations.

Even if the enzyme does not work under conditions demanded by industry, or for a broad range of substrates, all is not lost. Techniques such as protein engineering and directed evolution are increasingly being used to rapidly optimize the characteristics of enzymes (such as their substrate scope, stability and selectivity) to make them compatible with industrial processes[8]. Indeed, enzymes are the ultimate tunable catalysts, and will therefore surely be combined with many other chemical catalysts in the future. ■

Nicholas J. Turner *is at the School of Chemistry, University of Manchester, Manchester Institute of Biotechnology, Manchester M1 7DN, UK.*
e-mail: nicholas.turner@manchester.ac.uk

1. Rothenberg, G. *Catalysis: Concepts and Green Applications* 2nd edn (Wiley, 2017).
2. Litman, Z. C., Wang, Y., Zhao, H. & Hartwig, J. F. *Nature* **560**, 355–359 (2018).
3. Rudroff, F. *et al. Nature Catal.* **1**, 12–22 (2018).
4. Hönig, M., Sondermann, P., Turner, N. J. & Carreira, E. M. *Angew. Chem. Int. Edn* **56**, 8942–8973 (2017).
5. Metternich, J. B. & Gilmour, R. *J. Am. Chem. Soc.* **137**, 11254–11257 (2015).
6. Guo, X., Okamoto, Y., Schreier, M. R., Ward, T. R. & Wenger, O. S. *Chem. Sci.* **9**, 5052–5056 (2018).
7. Lauder, K. *et al. Angew. Chem. Int. Edn* **57**, 5803–5807 (2018).
8. Turner, N. J. *Nature Chem. Biol.* **5**, 567–573 (2009).

VISION

Birds perceive colours in categories

Humans perceive colours in categories such as red, even though we can discern red hues including ruby and crimson. It emerges that birds also categorize colours and this affects their colour-discrimination ability. SEE LETTER P.365

ALMUT KELBER

The amount of information reaching our sensory organs every second would be overwhelming if it were not for our ability to categorize it. Colour perception is a good example of this phenomenon. When we pick strawberries, we can easily discriminate between unripe fruit and fruit of the many different shades of red that indicate ripeness. Caves *et al.*[1] report on page 365 that zebra finches (*Taeniopygia guttata*) can also perceive a continuum of colours as belonging to distinct categories, a phenomenon that affects birds' ability to distinguish similar colours.

Although we can easily discriminate between the different shades of ripe strawberries, we tend to generalize and treat these shades as being equivalent. When comparing colours, if the differences between them are on the same scale of separation, our ability to perceive differences between colours from two separate categories, say 'red' and 'orange', is enhanced compared with our ability to perceive differences in colours that are both within one of these categories[2,3]. This enhanced ability to distinguish between colours if the colours are in separate categories is called categorical colour perception.

The preconditions necessary for the ability to perceive colours in distinct categories had already been demonstrated in birds. Humans and our close relatives have evolved to have three types of colour-sensing cone cell in the eye, and birds have evolved to have four types[4,5]. Birds have impressive colour-discrimination abilities[4], including the capacity to perceive the ultraviolet range of the spectrum. A remarkable earlier study[5] provided clear evidence that birds can generalize among certain colours, and thus divide the continuum of the colours that they perceive into discrete categories. But it was not known whether this ability affects how birds perceive similar colours and whether it helps them to spot key colour differences. Caves and colleagues investigated whether birds' ability to categorize colours affects their colour-discrimination abilities, and thus whether these animals have categorical colour perception.

"Birds are the only animals, besides primates, in which categorical colour perception has now been demonstrated."

The authors created an ingenious experimental set-up. Female zebra finches were presented with a device in which food was hidden beneath coloured discs. Food was present beneath bicoloured discs and absent below discs composed of a single colour. This training scheme allowed the authors to test how well the birds recognized colour differences by their ability to identify bicoloured discs when searching for food.

The authors studied a range of colours from orange to red, evenly dividing this part of the spectrum into eight shades of colour. Caves and colleagues made great efforts, using physiological models of bird colour vision, to make all of the steps between the shades equivalently

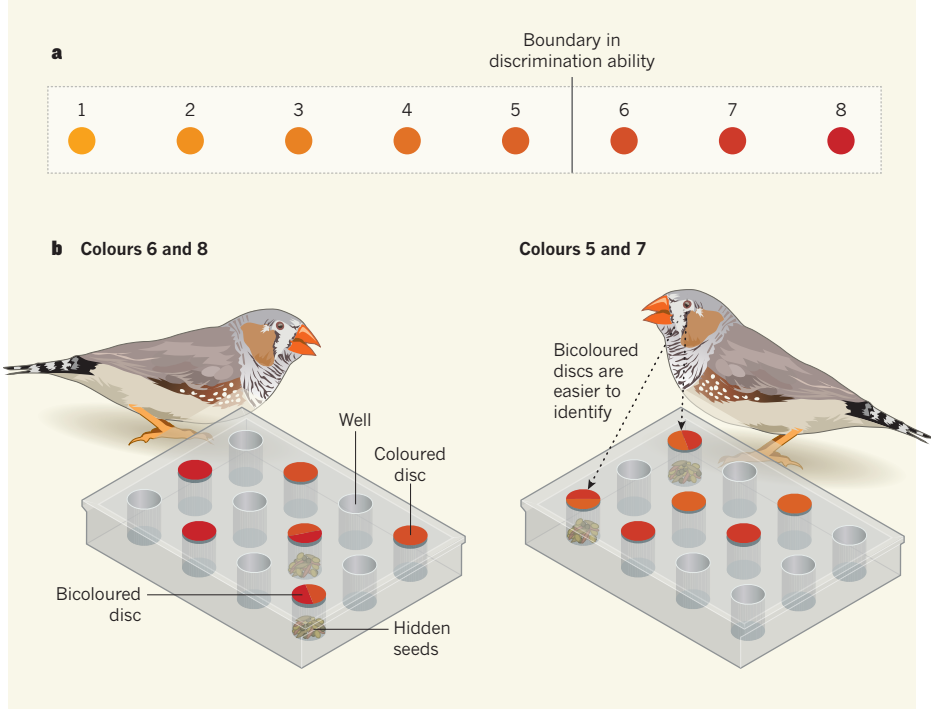


Figure 1 | Categorical colour perception in birds. Caves *et al.*[1] investigated colour perception in zebra finches (*Taeniopygia guttata*). The authors found that, like humans, the birds group colours into categories and this affects their colour-discrimination ability — a phenomenon called categorical colour perception. **a**, The authors tested birds' ability to distinguish between neighbouring colour pairs in eight evenly separated shades on the orange to red spectrum. Birds were substantially better at discriminating between shades 5 and 6 than between other pairs, suggesting that this represents a colour-category boundary. **b**, To test whether birds have the capacity for categorical colour perception, the authors presented birds with a device that had been designed so that the ability to distinguish between two colours and identify bicoloured discs enabled birds to access a food reward. The wells of the device contained seeds covered with bicoloured discs and empty wells that were either uncovered or covered by single-coloured discs of each of the colours on the bicoloured discs. The finches were less successful at identifying bicoloured discs if the colours were on the same side of the 5/6 category boundary (such as colours 6 and 8) than if the colours were in different categories (5 and 7).

sized on the birds' colour scale. These colours are worthy of attention because the zebra finch beak is red or orange. Beak colour depends on the amount of astaxanthin pigment deposited, which reflects the health of an individual's immune system[4], hence these colours might provide information about an individual's fitness. Females seem to be able to discriminate not only between males that have red or orange beaks, but also between males that have beaks of differing red shades[6]. However, whether female preference for males depends on male beak shade is debated[6].

Caves and colleagues first tested the finches using neighbouring pairs of shades from their eight-step colour scale and observed that birds distinguished between two of the shades better than between any other pair of neighbouring hues. This suggests that a putative border is present between the red and orange shades. The authors then investigated whether the birds were better at discriminating between pairs of colours of a similar level of shade separation that cross the proposed category boundary, compared with their ability to discriminate between colour pairs from one side of the category boundary (Fig. 1). Zebra finches passed this key test, demonstrating their capacity for categorical colour perception.

This result is fascinating and thought-provoking for many reasons. Birds are the only animals, besides primates[2,3], in which categorical colour perception has now been demonstrated. More work should be done to investigate whether other aspects of colour, such as intensity and spectral purity, influence categorical perception in birds. It would also be interesting to determine whether zebra finches' ability to group colours into 'red' and 'orange' has relevance for mate choice. However, this could be difficult to test because mate selection might depend on a range of male characteristics, such as the rate of male courtship displays[7], rather than only beak colour.

The work also has implications for our understanding of human colour perception. There is an ongoing debate about whether language — including colour terms such as red, blue, green and yellow — influences colour perception. One school of thought holds that colour categories have a cultural and linguistic basis[2]. The hallmark of categorical perception — faster and more-accurate discrimination of colours in different colour categories — is seen only if a subject's language has names for the specific colour categories being compared[2].

The other school of thought contends that colour perception has a biological basis that is not dependent on cultural and linguistic influences. Evidence to support this viewpoint includes the observation that terms for specific colours cluster around the same hues across different languages[3], and the fact that infants can discriminate between red, green, blue, yellow and purple before they have learnt the words for these colours[3]. Caves and colleagues' finding that birds have the capacity for categorical colour perception adds more evidence to support the biological basis of this phenomenon.

Why might categorization be important, and how does it fit into the broader context of signal perception? The term 'categorical perception' was coined to describe the human ability to distinguish sounds in discrete units, called phonemes, that help to discriminate one word from another (such as the sounds 'd', 't', 'b' and 'p' in the English words bad, bat, pad and pat[8]). Perception of phoneme-like elements also occurs in other animals, including birds[9]. Categorical perception could be described as a top-down mechanism to focus on key sensory cues by separating such signals from the enormous volume of irrelevant information. Another way to achieve this separation is a bottom-up approach termed 'matched filter', a concept which proposes that many animals' sensory organs are designed as filters that perceive only the range of information that is relevant to the organism[10]. These two approaches could together enable animals to handle the vast amount of sensory input that is needed to inform their choices and behaviours.

The level of contribution of these processes, and how they evolved in different animal clades, are topics worthy of further study. Caves and colleagues' work on zebra finches might be the start of a wider survey of categorical perception of colour in other animals. ■

Almut Kelber *is in the Department of Biology, Lund University, 22362 Lund, Sweden.*
e-mail: almut.kelber@biol.lu.se

1. Caves, E. M. *et al. Nature* **560**, 365–367 (2018).
2. Roberson, D., Davies, I. & Davidoff, J. *J. Exp. Psychol. Gen.* **129**, 369–398 (2000).
3. Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M. & Franklin, A. *Proc. Natl Acad. Sci. USA* **114**, 5545–5550 (2017).
4. Olsson, P., Lind, O. & Kelber, A. *J. Exp. Biol.* **218**, 184–193 (2015).
5. Jones, C. D., Osorio, D. & Baddeley, R. J. *Proc. R. Soc. B* **268**, 2077–2084 (2001).
6. Blount, J. D., Metcalfe, N. B., Birkhead, T. R. & Surai, P. F. *Science* **300**, 125–127 (2003).
7. Collins, S. A. & Ten Cate, C. *Anim. Behav.* **52**, 105–112 (1996).
8. Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. *J. Exp. Psychol.* **54**, 358–368 (1957).
9. Nelson, D. A. & Marler, P. *Science* **244**, 976–978 (1989).
10. Wehner, R. *J. Comp. Physiol. A* **161**, 511–531 (1987).

This article was published online on 1 August 2018.

GENOMICS

Profile of an unknown airway cell

RNA sequencing of single cells in the mammalian trachea reveals a previously unknown airway cell that expresses genes involved in fluid and solute balance, and that might play a part in cystic fibrosis. SEE ARTICLE P.319 & LETTER P.377

KYLE J. TRAVAGLINI & MARK A. KRASNOW

Over the past two centuries, light and electron microscopy have enabled scientists to identify six or more cell types that line the mammalian airways, transporting oxygen into the body and protecting us from microbes and particles in the air we breathe. In the past two decades, the identification of molecular markers for most of these cell types has provided insight into their functions, and given researchers and physicians a way to locate and characterize the cells in clinical specimens. Montoro *et al.*[1] (page 319) and Plasschaert *et al.*[2] (page 377) now describe the full gene-expression profiles of cell types that line the mammalian trachea. In doing so, they have discovered a previously unknown cell type that could hold the key to understanding the disease cystic fibrosis.

The genomics revolution was spurred by the development of techniques to measure messenger RNA levels for every gene[3,4] and thus to profile gene expression across the genome. However, these early approaches required millions of cells to generate enough mRNA for analysis. As such, they were limited to producing a profile of average gene expression among all the cells in a population analysed.

It is only in the past decade that more-sensitive methods have become available to profile individual cells[5]. In the past couple of years, improvements in cell isolation and in mRNA capture, amplification and sequencing have allowed single-cell mRNA sequencing (scRNAseq) to be broadly applied across biology[6]. This technique has begun to reveal the diversity of cells in developing, mature and diseased tissues, without any need for prior knowledge of the cells or the ability to purify individual cell types[7]. Many groups are now reporting gene-expression profiles for thousands or tens of thousands of individual cells. These profiles can be used to reveal the gene-expression programs that govern dynamic biological processes such as cell differentiation[7], and to create molecular cell atlases of entire organs and even whole organisms[8–10].

In the current studies, both groups used scRNAseq to analyse tens of thousands of cells from the lining (epithelium) of the trachea of mice, and Plasschaert *et al.* also analysed human airway cells that had proliferated and differentiated in culture. Their analyses confirmed gene-expression profiles for the two most common cell types: club cells, which secrete components of the mucus that lines the airways, including antimicrobial and immune modulatory proteins; and ciliated cells, which carry protruding structures called cilia that swirl to clear mucus and debris.

In addition, the large number of cells analysed revealed expression profiles for some rare and less-well-characterized cell types: goblet cells, which produce mucus proteins; tuft cells, which are thought to act as immune sensors[11]; and neuroendocrine cells, which sense oxygen levels, irritants and stretch in the airways, and signal to other lung cells and the central nervous system. The analyses also uncovered molecularly distinct subpopulations of club, goblet and tuft cells.

The two studies then established the gene-expression profile of basal cells, which are located below the other cells in the epithelium (Fig. 1) and function as stem and progenitor cells. Montoro *et al.* combined scRNAseq with a technique to track basal cells, showing that basal cells can directly give rise not only to club cells, as previously shown[12], but also to all the minor cell types analysed except goblet cells. Goblet cells, like ciliated cells[12], seem to arise from club cells. The groups identified previously unknown molecular markers for each cell type, along with type-specific combinations of transcription factors, which are presumably needed to select and maintain each cell type's distinct properties during airway development and renewal.

The groups' most surprising and important finding was the discovery of a previously unknown cell type. The authors dubbed these rare cells pulmonary ionocytes, because the cells' gene-expression profile overlaps with that of ionocytes found in fish gills. In fish, these cells maintain normal solute concentrations by regulating the exchange of sodium, chloride and calcium ions between the animals' tissues and the surrounding water[13]. It is not yet clear whether pulmonary ionocytes serve a similar function in mammalian airways, although the cells do express multiple ion-transport genes.

In addition, both the fish and the mammalian cells produce a transcription factor of the Foxi family. In fish, the Foxi protein is required for fish cells to adopt the characteristics of ionocytes. Likewise, Montoro *et al.* found that the *Foxi1* gene is necessary for the expression of ionocyte markers in the mouse trachea, and Plasschaert *et al.* showed that the FOXI1 protein governs ionocyte identity in cultured human airway cells. Notably, both groups report that Foxi1 controls the expression of the gene *cystic fibrosis transmembrane conductance regulator* (*CFTR*) in pulmonary ionocytes.

The CFTR protein transports chloride ions

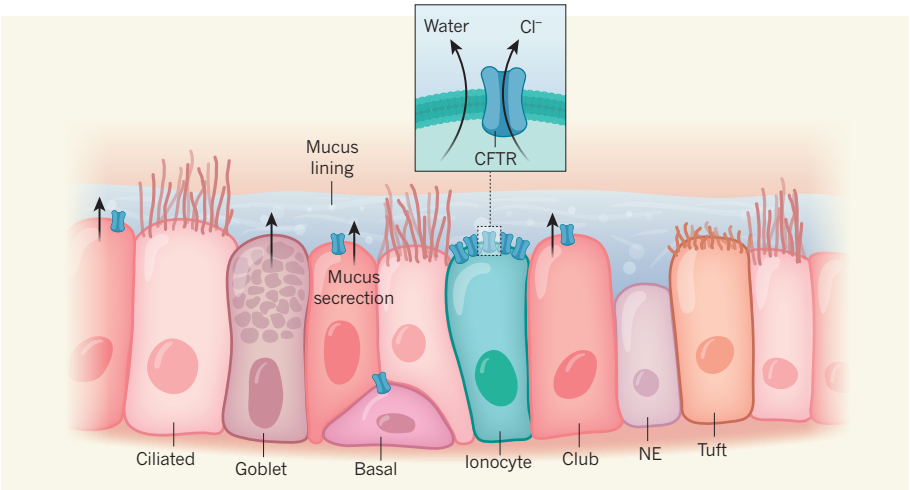


Figure 1 | Cells lining the trachea in mice. The inner surface of the trachea harbours multiple cell types. A protective lining of mucus is secreted by abundant club cells, rare goblet cells and submucosal glands (not shown). Ciliated cells bearing protrusions called cilia slowly propel mucus out of the lung. The surface also contains rare sensory neuroendocrine (NE) cells, tuft cells and basal progenitor cells. Two studies[1,2] have now identified another rare cell type on the airway surface: the ionocyte. Ionocytes highly express the gene *Cftr*, which encodes the CFTR ion channel, through which chloride ions (Cl^-) pass from the cell into the mucus, followed by water passage through a different channel (not shown). Basal cells and club cells also express *Cftr*, but at much lower levels than ionocytes. In people with the disease cystic fibrosis, CFTR is missing or defective, leading to thickening of the mucus, clogging of the airway, and repeated infections.

out of airway cells, causing water to flow out and so thinning the airway's lining of mucus (Fig. 1). When CFTR is absent or inactive, as in people who have cystic fibrosis, the mucus thickens and accumulates, causing airway obstruction and repeated infections and inflammation[14]. Determining which lung cells express *CFTR* and are directly affected in people with cystic fibrosis has been difficult because expression of this gene seems to be complex and variable along the airways[15]. The current papers demonstrate that the gene's expression is not as random as it had seemed: the bulk of the *CFTR* mRNA detected was from the rare pulmonary ionocytes, each of which highly express the gene. People with cystic fibrosis can also experience gastrointestinal symptoms and infertility. Perhaps *CFTR*-expressing ionocytes will be discovered in organs involved in these problems, too.

Mice harbouring mutations in *Cftr* do not develop cystic fibrosis — a curious fact that has long hampered research into the disease. Montoro *et al.* found that cultured airway epithelial cells generated from *Foxi1*-mutant mice have low *Cftr* expression but, paradoxically, higher than normal Cftr activity. Differences in pulmonary ionocytes between mice and humans, such as compensatory expression of another chloride channel when *Cftr* expression is lost in mice, might explain both this paradox and why *Cftr*-mutant mice do not model the disease.

Although these results suggest that ionocytes have a key role in airway biology and cystic fibrosis, much work is still needed to define their physiological functions, the role of CFTR in these functions, and how loss of CFTR causes or contributes to disease symptoms. Developing methods to genetically and pharmacologically manipulate ionocytes or replace them in model systems, and ultimately in patients, is another priority.

These papers provide excellent examples of how scRNAseq can transform long-established views of a tissue and a human disease. As scRNAseq tools improve and costs continue to drop, we will probably soon witness something similar for many human organs and diseases. ■

Kyle J. Travaglini *and* **Mark A. Krasnow**
are in the Department of Biochemistry and Howard Hughes Medical Institute, Stanford, California 94305, USA.
e-mail: krasnow@stanford.edu

1. Montoro, D. T. *et al. Nature* **560**, 319–324 (2018).
2. Plasschaert, L. W. *et al. Nature* **560**, 377–381 (2018).
3. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. *Science* **270**, 467–470 (1995).
4. Wang, Z., Gerstein, M. & Snyder, M. *Nature Rev. Genet.* **10**, 57–63 (2009).
5. Tang, F. *et al. Nature Methods* **6**, 377–382 (2009).
6. Wu, A. R. *et al. Nature Methods* **11**, 41–46 (2014).
7. Treutlein, B. *et al. Nature* **509**, 371–375 (2014).
8. Macosko, E. Z. *et al. Cell* **161**, 1202–1214 (2015).
9. The Tabula Muris Consortium. Preprint at bioRxiv https://dx.doi.org/10.1101/237446 (2017).
10. Han, X. *et al. Cell* **172**, 1091–1107 (2018).
11. Howitt, M. R. *et al. Science* **351**, 1329–1333 (2016).
12. Rock, J. R., Rawlins, E. L., Onaitis, M. W. & Hogan, B. L. *Dev. Biol.* **331**, 503 (2009).
13. Dymowska, A. K., Hwang, P.-P. & Goss, G. G. *Respir. Physiol. Neurobiol.* **184**, 282–292 (2012).
14. Elborn, J. S. *Lancet* **388**, 2519–2531 (2016).
15. Engelhardt, J. F., Zepeda, M., Cohn, J. A., Yankaskas, J. R. & Wilson, J. M. *J. Clin. Invest.* **93**, 737–749 (1994).

This article was published online on 1 August 2018.

ORGANIC CHEMISTRY

A practical route to 3D molecular diversity

Cycloaddition reactions are powerful tools for synthesizing three-dimensional molecules, but their scope has been limited. A creative solution to this problem opens up opportunities for drug discovery. SEE LETTER P.350

WENBO YE & ANG LI

Reactions known as cycloadditions are unparalleled in their ability to construct ring-containing molecules in a way that precisely controls the geometric arrangement of groups attached to the carbon atoms in the molecules — that is, the reactions offer great stereoselectivity. The power of these reactions has been demonstrated in numerous syntheses of complex natural products[1,2]. However, the scope of cycloadditions is limited to certain combinations of starting materials, which has restricted their use for making libraries of compounds in drug-discovery programs[3]. On page 350, Chen *et al.*[4] report a strategy that combines cycloadditions with another type of reaction, known as carbon–carbon cross coupling, to enable the modular and programmable preparation of cycloaddition-derived molecules.

Carbon–carbon (C–C) cross-coupling reactions are often used to form bonds between carbon atoms that are already part of a carbon–carbon double bond; such carbon atoms are said to have sp^2 orbital hybridization. Cross couplings between sp^2 carbons lend themselves to the modular synthetic routes

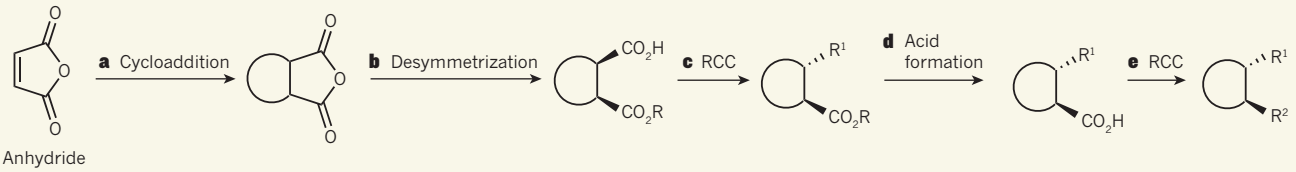


Figure 1 | A powerful strategy for making 3D molecules. Chen *et al.*[4] report a practically simple procedure that allows access to products that cannot be made easily and directly using reactions known as cycloadditions. **a**, An anhydride starting material undergoes a cycloaddition to form a 3D scaffold. The ring represents several different ring structures. **b**, A desymmetrization reaction then forms an intermediate that contains a carboxylic acid (CO_2H), in which the absolute stereochemistry (the geometric arrangement of groups) is fixed. Solid wedge bonds project above the plane of the page; R is methyl, benzyl or $CH_2CH_2Si(CH_3)_3$, and Si is silicon. **c**, The acid is replaced by a molecular appendage (R^1) in a radical cross-coupling (RCC) reaction. Broken wedges project below the plane of the page. **d**, **e**, A second acid group is generated (**d**), and is converted to a different appendage (R^2) in another RCC reaction (**e**).

used to make compound libraries. This has resulted in the predomination of sp^2-rich structures — which tend to be two-dimensional — in compounds tested for drug discovery. But 3D molecular structures can interact with biological targets in different ways from 2D ones, so are just as important for drug discovery. Three-dimensional molecules tend to be rich in sp^3 carbons, which have a different orbital hybridization from sp^2 carbons, and have the capacity to form four single bonds.

A cycloaddition known as the Diels–Alder reaction provides one of the most efficient means of building sp^3-rich ring systems. However, to achieve high reaction yields, stereoselectivity and regioselectivity (a preference to react at particular atoms) in Diels–Alder reactions, the electronic properties of the reactants should match. For example, in conventional Diels–Alder reactions, one of the reactants (known as a diene) should be electron-rich, whereas the other (the dienophile) should be electron-poor. This drastically reduces the number and diversity of Diels–Alder products that can be made. Chemistry students excitedly studying Diels–Alder reactions are often frustrated when they realize the restrictions involved. The constraints also prevent these reactions from being used in modular synthetic routes. Diels–Alder reactions have therefore been used much less often for drug discovery than have cross-coupling reactions[3], and so their advantages for controlling the stereochemistry (the geometric arrangement of groups) of ring-containing molecules have not been fully exploited by medicinal chemists.

Workers from the same group as Chen *et al.* previously developed a type of cross-coupling reaction known as radical cross-coupling (RCC)[5–9]. This is a useful tool for converting carboxylic acids (compounds that contain CO_2H groups) into products that contain alkyl, alkenyl, alkynyl or aryl groups (hydrocarbon groups that represent all the possible bonding geometries of carbon atoms). In the current paper, Chen and colleagues use carboxylic acids as a link that allows them to combine RCC and Diels–Alder reactions. The connection can be made because compounds known as anhydrides and esters serve as electronically favourable dienophiles in Diels–Alder reactions, and can then be converted into acids to take part in various RCC reactions.

On the basis of this strategy, the authors devised a simple, modular, five-step sequence to generate molecules that have 3D structural complexity (Fig. 1). In the first step, a Diels–Alder reaction involving an anhydride or ester builds a 3D molecular scaffold. This is followed by a 'desymmetrization' reaction[10], which generates a carboxylic acid and sets the absolute stereochemistry in the resulting product. In the third step, an RCC reaction replaces the acid group with a molecular appendage, producing an intermediate that is then hydrolysed to generate a second carboxylic acid. This is used in the final step: another RCC reaction, which introduces a second appendage.

The two RCC steps allow Diels–Alder-type products to be made that couldn't be synthesized directly in a Diels–Alder reaction because the starting materials would have mismatched electronic properties. Another remarkable feature of this sequence is the clever mechanism of stereocontrol: the first appendage is attached to the molecular scaffold at an orientation that is governed by a nearby group produced during the desymmetrization reaction, and the orientation of the second appendage is governed by the orientation of the first. The final product therefore comprises mostly one isomer in which the two appendages are fixed in what is known as a *trans* orientation to each other.

Chen *et al.* went on to extend this chemistry from Diels–Alder reactions to three other types of cycloaddition reaction t hat construct rings formed of three, four or five atoms. Moreover, one of the RCC steps could be replaced with a reaction that allowed the formation of a carbon–nitrogen bond, rather than a C–C bond. The researchers suggest that bonds from carbon to other types of atom could also be made, to produce an even more structurally diverse set of final products.

The authors showcase their chemistry by using it to make natural products, pharmaceuticals and key intermediates used in the synthesis of such compounds. An example is the antipsychotic drug asenapine, which is usually manufactured as a mixture of two mirror-image isomers (enantiomers). One of these is more effective as a drug than the other, but is difficult to synthesize as a single enantiomer. Starting from a symmetrical anhydride molecule, Chen *et al.* show that their strategy can be used to make this enantiomer in a practically simple, short synthetic route and in good overall yield, offering several advantages over the previously reported synthesis[11].

Chen and colleagues' chemistry is easy to carry out in the laboratory. Our frustrated chemistry students can therefore immediately start to synthesize compounds that fall outside the conventional scope of Diels–Alder reactions described in their textbooks. All they have to do is to collect the anhydrides reported by the authors, the organic catalyst needed for the desymmetrization, the metal catalysts and the activating reagent used for RCC, and various other commercially available molecular building blocks (such as boronic acids). Some of these items will be used in every reaction sequence, and so the task would become even easier if a chemical-supplies company could market a kit that contained these items.

Chen and colleagues' work will redefine how chemists think about synthesizing Diels–Alder-type products, and will find numerous applications for synthesizing molecules rich in sp^3 carbons for drug discovery. The remaining challenges mainly concern the limitations of RCC. For example, the method can currently produce only a *trans* arrangement of the two appendages introduced in the RCC steps; to make the other arrangement, the stereochemical outcome of the second RCC reaction would need to be changed. If such issues can be addressed, then Chen and colleagues' reactions will become even more powerful tools for synthesis than they already are. ■

Wenbo Ye *and* **Ang Li** *are at the State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China.*
e-mail: ali@sioc.ac.cn

1. Nishiwaki, N. (ed.) *Methods and Applications of Cycloaddition Reactions in Organic Syntheses* (Wiley, 2014).
2. Nicolaou, K. C., Snyder, S. A., Montagnon, T. &

Vassilikogiannakis, G. *Angew. Chem. Int. Edn* **41**, 1668–1698 (2002).
3. Brown, D. G. & Boström, J. *J. Med. Chem.* **59**, 4443–4458 (2016).
4. Chen, T.-G. *et al. Nature* **560**, 350–354 (2018).
5. Cornella, J. *et al. J. Am. Chem. Soc.* **138**, 2174–2177 (2016).
6. Qin, T. *et al. Science* **352**, 801–805 (2016).
7. Wang, J. *et al. Angew. Chem. Int. Edn* **55**, 9676–9679 (2016).
8. Edwards, J. T. *et al. Nature* **545**, 213–218 (2017).
9. Smith, J. M. *et al. Angew. Chem. Int. Edn* **56**, 11906–11910 (2017).
10. Chen, Y., Tian, S.-K. & Deng, L. *J. Am. Chem. Soc.* **122**, 9542–9543 (2000).
11. Anugu, R. R., Mainkar, P. S., Sridhar, B. & Chandrasekhar, S. *Org. Biomol. Chem.* **14**, 1332–1337 (2016).

This article was published online on 6 August 2018.

50 Years Ago

Whatever advantages the distant future may offer from the manned exploration of space, it is clear that the present benefits of the space race are few and far between … A fortnight's conference is being held … under the title "Space Science and Technology —Benefits to Developing Countries". The introductory pamphlet … describes in a style of sustained optimism the cornucopia of technological blessings which future satellite systems will rain down on the poor and rich alike. In the near future, reflector satellites will shed light on the night earth, the pamphlet says, and "by providing illumination for construction, lumbering, fishing and other outdoor industries, could conceivably have an important effect on the economic growth of the developing nations". The pamphlet does not discuss the catastrophic effect of such a satellite on biological rhythms, nor does it explain in what manner night-time fishing and lumbering will boost any nation's economy.
From *Nature* 17 August 1968

100 Years Ago

The entrance of the United States of America into the war has prompted Mr. A. Hansen to write to *Science* pointing out that the States possess no national floral emblem. France has its fleur-de-lis, England the rose, Scotland the thistle, but America has no flower with which it is associated in people's minds. Mr. Hansen points out the various characteristics required for a national flower, and comes to the conclusion that the columbine, which is in flower from April to July, is probably the most suitable for the purpose. The correspondence of the generic name Aquilegia with the Latin name of the eagle is also considered to be a point in its favour.
From *Nature* 15 August 1918

PARTICLE PHYSICS

Special relativity validated by neutrinos

Neutrinos are tiny, ghost-like particles that habitually change identity. A measurement of the rate of change in high-energy neutrinos racing through Earth provides a record-breaking test of Einstein's special theory of relativity.

MATTHEW MEWES

The existence of extremely light, electrically neutral particles called neutrinos was first postulated in 1930 to explain an apparent violation of energy conservation in the decays of certain unstable atomic nuclei. Writing in *Nature Physics*, the IceCube Collaboration[1] now uses neutrinos seen in the world's largest particle detector to scrutinize another cornerstone of physics: Lorentz invariance. This principle states that the laws of physics are independent of the speed and orientation of the experimenter's frame of reference, and serves as the mathematical foundation for Albert Einstein's special theory of relativity. Scouring their data for signs of broken Lorentz invariance, the authors carry out one of the most stringent tests of special relativity so far, and demonstrate how the peculiarities of neutrinos can be used to probe the foundations of modern physics.

Physicists generally assume that Lorentz invariance holds exactly. However, in the late 1990s, the principle began to be systematically challenged[2], largely because of the possibility that it was broken slightly in proposed theories of fundamental physics, such as string theory[3]. Over the past two decades, researchers have tested Lorentz invariance in objects ranging from photons to the Moon[4].

The IceCube Collaboration instead tested the principle using neutrinos. Neutrinos interact with matter through the weak force — one of the four fundamental forces of nature. The influence of the weak force is limited to minute distances. As a result, interactions between neutrinos and matter are extremely improbable, and a neutrino can easily traverse through the entire Earth unimpeded. This poses a challenge for physicists trying to study these elusive particles, because almost every neutrino will simply pass through any detector completely unnoticed.

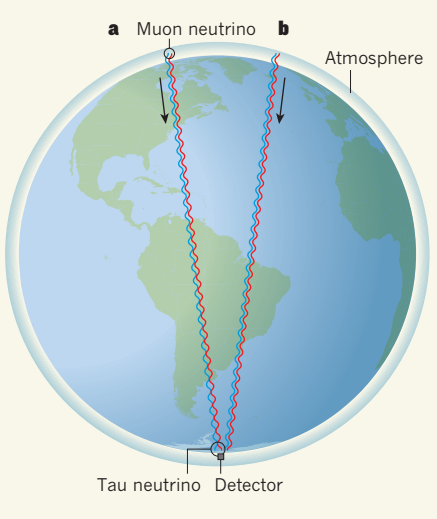


Figure 1 | Propagation of neutrinos through Earth. There are three known types of neutrino: electron, muon and tau. **a**, A muon neutrino produced in Earth's atmosphere can be thought of as the combination of two quantum-mechanical waves (red and blue) that are in phase — the peaks of the waves are observed at the same time. If a principle known as Lorentz invariance were violated, these waves could travel at different speeds through Earth's interior and be detected in the out-of-phase tau-neutrino state. **b**, The IceCube Collaboration[1] reports no evidence of such conversion, constraining the extent to which Lorentz invariance could be violated.

The IceCube Neutrino Observatory, located at the South Pole, remedies this problem by monitoring an immense target volume to glimpse the exceedingly rare interactions. At the heart of the detector are more than 5,000 light sensors, which are focused on 1 cubic kilometre (1 billion tonnes) of ice. The sensors constantly look for the telltale flashes of light that are produced when a neutrino collides with a particle in the ice.

The main goal of the IceCube Neutrino Observatory is to observe comparatively scarce neutrinos that are produced during some of the Universe's most violent astrophysical events. However, in its test of Lorentz invariance, the collaboration studied more-abundant neutrinos that are generated when fast-moving charged particles from space collide with atoms in Earth's atmosphere. There are three known types of neutrino: electron, muon and tau. Most of the neutrinos produced in the atmosphere are muon neutrinos.

Atmospheric neutrinos generated around the globe travel freely to the South Pole, but can change type along the way. Such changes stem from the fact that electron, muon and tau neutrinos are not particles in the usual sense. They are actually quantum combinations of three 'real' particles — ν_1, ν_2 and ν_3 — that have tiny but different masses.

In a simple approximation relevant to the IceCube experiment, the birth of a muon neutrino in the atmosphere can be thought of as the simultaneous production of two quantum-mechanical waves: one for ν_2 and one for ν_3 (Fig. 1). These waves are observed as a muon neutrino only because they are in phase, which means the peaks of the two waves are seen at the same time. By contrast, a tau neutrino results from out-of-phase waves, whereby the peak of one wave arrives with the valley of the other.

If neutrinos were massless and Lorentz invariance held exactly, the two waves would simply travel in unison, always maintaining the in-phase muon-neutrino state. However, small differences in the masses of ν_2 and ν_3 or broken Lorentz invariance could cause the waves to travel at slightly different speeds, leading to a gradual shift from the muon-neutrino state to the out-of-phase tau-neutrino state. Such transitions are known as neutrino oscillations and enable the IceCube detector to pick out potential violations of Lorentz invariance. Oscillations resulting from mass differences are expected to be negligible at the neutrino energies considered in the authors' analysis, so the observation of an oscillation would signal a possible breakdown of special relativity.

The IceCube Collaboration is not the first group to seek Lorentz-invariance violation in neutrino oscillations[5–10]. However, two key factors allowed the authors to carry out the most precise search so far. First, atmospheric neutrinos that are produced on the opposite side of Earth to the detector travel a large distance (almost 13,000 km) before being observed, maximizing the probability that a potential oscillation will occur. Second, the large size of the detector allows neutrinos to be observed that have much higher energies than those that can be seen in other experiments.

Such high energies imply that the quantum-mechanical waves have tiny wavelengths, down to less than one-billionth of the width of an atom. The IceCube Collaboration saw no sign of oscillations, and therefore inferred that the peaks of the waves associated with ν_2 and ν_3 are shifted by no more than this distance after travelling the diameter of Earth. Consequently, the speeds of the waves differ by no more than a few parts per 10^{28} — a result that is one of the most precise speed comparisons in history.

The authors' analysis provides support for special relativity and places tight constraints on a number of different classes of Lorentz-invariance violation, many for the first time. Although already impressive, the IceCube experiment has yet to reach its full potential. Because of limited data, the authors restricted their attention to violations that are independent of the direction of neutrino propagation, neglecting possible direction-dependent violations that could arise more generally.

With a greater number of neutrino detections, the experiment, or a larger future version[11], could search for direction-dependent violations. Eventually, similar studies involving more-energetic astrophysical neutrinos propagating over astronomical distances could test the foundations of physics at unprecedented levels. ■

Matthew Mewes *is in the Department of Physics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, California 93407, USA.*
e-mail: mmewes@calpoly.edu

1. The IceCube Collaboration. *Nature Phys.* https://doi.org/10.1038/s41567-018-0172-2 (2018).
2. Colladay, D. & Kostelecký, V. A. *Phys. Rev. D* **58**, 116002 (1998).
3. Kostelecký, V. A. & Potting, R. *Phys. Rev. D* **51**, 3923–3935 (1995).
4. Kostelecký, V. A. & Russell, N. Preprint at https://arxiv.org/abs/0801.0287v11 (2018).
5. LSND Collaboration. *Phys. Rev. D* **72**, 076004 (2005).
6. The MINOS Collaboration. *Phys. Rev. D* **85**, 031101 (2012).
7. Double Chooz Collaboration. *Phys. Rev. D* **86**, 112009 (2012).
8. MiniBooNE Collaboration. *Phys. Lett. B* **718**, 1303–1308 (2013).
9. Super-Kamiokande Collaboration. *Phys. Rev. D* **91**, 052003 (2015).
10. T2K Collaboration. *Phys. Rev. D* **95**, 111101 (2017).
11. IceCube–Gen2 Collaboration. Preprint at https://arXiv.org/abs/1412.5106 (2014).

GENOME EDITING

Chromosomes get together

Genome-editing approaches have been used to fuse 16 yeast chromosomes to produce yeast strains with only 1 or 2 chromosomes. Surprisingly, this fusion has little effect on cell fitness. SEE ARTICLE P.331 & LETTER P.392

GIANNI LITI

The genomes of nucleus-bearing organisms are divided into linear chromosomes. The number of chromosomes ranges from one to hundreds across species. But why is there such variation? Do specific chromosome numbers hold an advantage for particular species? Shao *et al.*[1] (page 331) and Luo *et al.*[2] (page 392) independently manipulate the genome of the budding yeast *Saccharomyces cerevisiae* by systematically fusing chromosomes, enabling the researchers to explore the consequences of chromosome-number reduction.

Normal *S. cerevisiae* genomes have 16 distinct chromosomes ($n = 16$), which range from 230 to 1,532 kilobases in length[3]. To function correctly, yeast chromosomes need protective structures called telomeres at both ends, and only one centromere — a region that ensures the accurate segregation of chromosomes into mother and daughter cells during cell division. Simply fusing the ends of two chromosomes is therefore not a viable strategy for reducing chromosome number because it would produce chromosomes containing two centromeres.

To solve this problem, the two groups used genome-editing tools to fuse sequences found adjacent to one of the telomeres in each chromosome, and to simultaneously remove one of the two centromeres (Fig. 1). Using this approach, they reduced the chromosome number step by step, producing strains that had progressively lower values of n. The fusion strains comprised genomic material that is almost identical to that of normal *S. cerevisiae*, differing only in chromosome number and by a few non-essential genes that were deleted during strain creation.

Luo *et al.* produced an $n = 2$ strain containing chromosomes that were each about 6,000 kb long. However, they were unable to fuse the two chromosomes into one as part of a viable cell. By contrast, Shao *et al.* successfully fused the entire *S. cerevisiae* genome into a single chromosome in a functional yeast.

Given that each group used similar strategies, it is interesting to consider why only one of the teams could fuse the final two chromosomes. A possible explanation is that

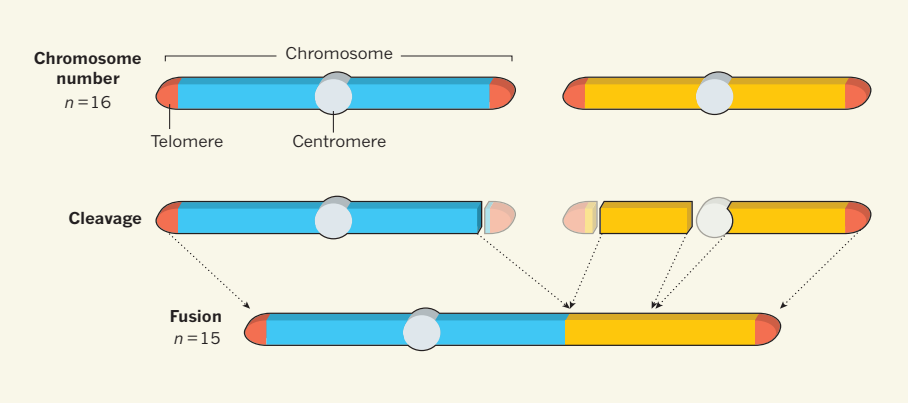


Figure 1 | Fusing chromosomes one by one. Two groups[1,2] have fused all 16 chromosomes ($n = 16$) of the budding yeast *Saccharomyces cerevisiae* to produce strains that have only one[1] or two[2] long chromosomes. In *S. cerevisiae*, each chromosome must have protective structures known as telomeres at both ends, as well as a single structure called a centromere that is essential for normal chromosome segregation during cell division. To generate fused chromosomes with this composition, the groups used genome-editing techniques to cleave sequences found next to one of the telomeres in each chromosome, and simultaneously removed one of the two centromeres (just one cleavage site is sufficient to trigger removal of the entire centromere). They then fused the cleaved portions. By repeating this process, they reduced the chromosome number in a stepwise manner, producing yeast cells that have progressively lower values of n.

the groups fused the yeast chromosomes in different orders and orientations. Perhaps such factors matter, which could mean that only certain final genome structures are attainable. In the future, reducing the chromosome number through a variety of fusion paths might reveal how chromosomal structures affect cell viability. Another possibility is that mutations introduced accidentally during the chromosome-fusion experiments affected cell tolerance to the new genome organization.

Both groups then investigated the biological implications of chromosome fusion. Overall, organismal traits such as cell growth, size and shape seem to be buffered throughout the series of fusions. Notably, the expression of only a few genes was altered considerably in either the $n = 2$ or $n = 1$ strains. Most of the observed increases in gene expression can be explained by there being fewer genes located near telomeres, which promote transcriptional silencing[4].

Such transcriptional stability is in contrast to the widespread transcriptional variation that is seen when yeast undergoes other types of chromosomal modification, such as inversions of particular regions[5]. Shao *et al.* show that this stability reflects the fact that there are only modest changes to the intrachromosomal interactions that usually take place, which can modulate gene expression. However, the interchromosomal-interaction landscape changes drastically, owing to the depletion of centromeres, which drive the 3D configuration of the yeast genome[6].

The yeast strains generated by the groups are haploid — they contain only one copy of each chromosome. Haploid yeast reproduce asexually, but they can also mate through sexual reproduction to form diploid yeast, which contain two copies of each chromosome. Diploid yeast can then divide through a process called meiosis to produce haploid spores that mature into haploid cells. The groups showed that the $n = 1$ and $n = 2$ strains can undergo sexual reproduction, albeit with reduced efficiency compared with wild-type yeast, and produce spores that are slightly less viable.

During meiosis, genetic material is exchanged between paired chromosomes in a process called recombination. Because the genomes of all cells from a given fusion strain are identical, they lack the genetic variability that researchers need to map recombination through the generations. As such, the two groups could not characterize how chromosome reduction affects recombination. The high spore viability of each fusion strain indicates that some recombination might occur, ensuring proper chromosome segregation. However, the greatly reduced chromosome number essentially eliminates any risk of mis-segregation.

Luo *et al.* mated strains that had different chromosome numbers, and then investigated spore viability and production in the resulting hybrid strains, to determine at what point the fusion strain could no longer produce viable spores (a phenomenon called reproductive isolation). As predicted[7], an increasing difference in chromosome number had an increasing effect on spore viability until, in hybrids generated by crossing haploid strains that have $n = 16$ and $n = 8$, none of the spores produced were viable. Moreover, spore production was arrested when the difference in chromosome number became any larger.

This is unexpected, especially given that diploid hybrids that are sterile because of high sequence divergence or differently arranged genomes between their two sets of chromosomes can progress efficiently through meiosis, despite producing inviable spores[7]. The mechanism that underlies the reproductive isolation seen by Luo and colleagues remains to be determined. Future work using synthetic genomes, which can be edited at the single-nucleotide level, will allow the introduction of genetic variants on both local and genome-wide scales, enabling the in-depth, systematic analysis of the factors that prevent species from breeding, as well as the genomic changes that prompt reproductive isolation.

Both studies concluded that reduced chromosome number causes no major growth defects when cells are grown under various conditions and stresses. Small fitness defects were most evident in the $n = 1$ strain, consistent with the fact that this chromosome configuration is challenging to obtain. Although these fitness differences seem mild in a laboratory setting, they could become more harmful in the natural environment. Indeed, Shao and colleagues' $n = 1$ strain was quickly outcompeted by a normal strain of *S. cerevisiae* when the two were cultured together. This is consistent with the idea that the structure of *S. cerevisiae* chromosomes has remained highly stable for several million years[8,9], although reductions in chromosome number through telomere fusion and centromere loss occurred repeatedly over longer evolutionary timescales[10].

The short generation time of *S. cerevisiae* means that, in the future, the evolution of strains that have a reduced chromosome number could be tracked in the lab, in long-term experiments that run for months or years. Such experiments will enable researchers to map adaptive changes that restore fitness in strains that have a reduced number of chromosomes, and to accurately measure genome stability in these yeast.

Beyond the current findings, these engineered yeast strains constitute powerful resources for studying fundamental concepts in chromosome biology, including replication, recombination and segregation. The chromosome-engineering approach might also be applicable to organisms that have more-complex genomes. However, the presence of highly complex DNA sequences in the regions that surround telomeres and centromeres in these organisms will make this a challenging task. ■

Gianni Liti *is at the University of Côte d'Azur, CNRS, INSERM, IRCAN, 06107 Nice, France.*
e-mail: gianni.liti@unice.fr

1. Shao, Y. *et al. Nature* **560**, 331–335 (2018).
2. Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. *Nature* **560**, 392–396 (2018).
3. Goffeau, A. *et al. Science* **274**, 546–567 (1996).
4. Pryde, F. E. & Louis, E. J. *EMBO J.* **18**, 2538–2550 (1999).
5. Naseeb, S. *et al. Mol. Biol. Evol.* **33**, 1679–1696 (2016).
6. Duan, Z. *et al. Nature* **465**, 363–367 (2010).
7. Greig, D. *Heredity* **102**, 39–44 (2009).
8. Yue, J.-X. *et al. Nature Genet.* **49**, 913–924 (2017).
9. Liti, G., Barton, D. B. & Louis, E. J. *Genetics* **174**, 839–850 (2006).
10. Gordon, J. L., Byrne, K. P. & Wolfe, K. H. *PLoS Genet.* **7**, e1002190 (2011).

This article was published online on 1 August 2018.

ARTICLE

https://doi.org/10.1038/s41586-018-0393-7

A revised airway epithelial hierarchy includes CFTR-expressing ionocytes

Daniel T. Montoro[1,2,3,4,20], Adam L. Haber[4,20], Moshe Biton[4,5,20], Vladimir Vinarsky[1,2,3], Brian Lin[1,2,3], Susan E. Birket[6,7], Feng Yuan[8], Sijia Chen[9], Hui Min Leung[10,11], Jorge Villoria[1,2,3], Noga Rogel[4], Grace Burgin[4], Alexander M. Tsankov[4], Avinash Waghray[1,2,3,4], Michal Slyper[4], Julia Waldman[4], Lan Nguyen[4], Danielle Dionne[4], Orit Rozenblatt-Rosen[4], Purushothama Rao Tata[12,13,14,15], Hongmei Mou[16,17], Manjunatha Shivaraju[1,2,3], Hermann Bihler[18], Martin Mense[18], Guillermo J. Tearney[10,11], Steven M. Rowe[6,7], John F. Engelhardt[8], Aviv Regev[4,19]* & Jayaraj Rajagopal[1,2,3,4]*

The airways of the lung are the primary sites of disease in asthma and cystic fibrosis. Here we study the cellular composition and hierarchy of the mouse tracheal epithelium by single-cell RNA-sequencing (scRNA-seq) and in vivo lineage tracing. We identify a rare cell type, the Foxi1$^+$ pulmonary ionocyte; functional variations in club cells based on their location; a distinct cell type in high turnover squamous epithelial structures that we term 'hillocks'; and disease-relevant subsets of tuft and goblet cells. We developed 'pulse-seq', combining scRNA-seq and lineage tracing, to show that tuft, neuroendocrine and ionocyte cells are continually and directly replenished by basal progenitor cells. Ionocytes are the major source of transcripts of the cystic fibrosis transmembrane conductance regulator in both mouse (*Cftr*) and human (*CFTR*). Knockout of *Foxi1* in mouse ionocytes causes loss of *Cftr* expression and disrupts airway fluid and mucus physiology, phenotypes that are characteristic of cystic fibrosis. By associating cell-type-specific expression programs with key disease genes, we establish a new cellular narrative for airways disease.

The airways conduct oxygen from the atmosphere to the distal gas-exchanging alveoli and are the loci of major diseases, including asthma, chronic obstructive pulmonary disease and cystic fibrosis. The predominant airway epithelial cell types include basal progenitor cells, secretory club cells and ciliated cells[1]. Rare cell types such as solitary neuroendocrine, goblet and tuft cells have received less scrutiny, and their lineage relationships and functions remain poorly understood. Of note, diseases of the airway occur at distinct proximodistal sites along the respiratory tree. This phenomenon has been attributed to physical factors governing the localized deposition of inhaled particulates, toxins, smoke and allergens[2]. Whether disease heterogeneity also reflects cellular heterogeneity, which varies along the airway tree, is unknown. scRNA-seq studies[3–6] have begun to delineate cell type diversity and lineage hierarchy in the lung.

Here we combine massively parallel scRNA-seq (also performed in the accompanying Letter[7]) and in vivo lineage tracing in the adult mouse tracheal epithelium. The resulting finer taxonomy highlights new cell types and subtypes, reveals new tissue structures and refines lineage relations. These findings reframe our understanding of both Mendelian and complex multigenic airway diseases, including cystic fibrosis and asthma.

scRNA-seq reveals new disease-associated cell types

We initially profiled 7,494 EpCAM$^+$ tracheal epithelial cells from C57BL/6 wild-type mice ($n = 4$) and *Foxj1*-EGFP ciliated cell reporter mice ($n = 2$), using complementary scRNA-seq approaches: massively parallel droplet-based 3′ scRNA-seq ($k = 7$, 193 cells) and full-length scRNA-seq ($k = 301$ cells; Fig. 1a, Extended Data Figs. 1, 2). We partitioned the cells profiled by 3′ scRNA-seq into seven distinct clusters annotated post hoc by expression of known marker genes (Fig. 1b, Extended Data Fig. 1). Each cluster mapped to known abundant (basal, club, ciliated) or rare (tuft, neuroendocrine, goblet) epithelial cell types, except for one cluster (Fig. 1b) that contained cells with expression profiles similar to those of ionocytes found in the skin of *Xenopus* and zebrafish[8,9]. We also recovered matching clusters using full-length scRNA-seq of 301 EpCAM$^+$CD45$^-$ epithelial cells from proximal and distal tracheal segments of C57BL/6 wild-type mice, with the exception of goblet cells ($n = 3$; Fig. 1a, Extended Data Figs. 2, 3a, b).

We identified new consensus markers (Extended Data Figs. 1f, 3b, Supplementary Tables 1–3) and cell-type-specific transcription factors (false discovery rate, FDR < 0.01, likelihood-ratio test (LRT)), Extended Data Fig. 3c, Supplementary Table 4). To our knowledge, *Nfia* is the first transcription factor that is known to be enriched in club cells. *Nfia* regulates Notch signalling, which is known to be required for club cell maintenance[10,11] *Ascl1*, *Ascl2* and *Ascl3*, also associated with Notch signalling[12,13], are enriched in the rare solitary neuroendocrine cells, tufts cells and ionocytes, respectively (FDR < 0.0001, LRT). Goblet cells specifically express *Foxq1*, which is essential for mucin expression in gastric epithelia[14].

Some cell-type-specific markers, including *Cdhr3* (ciliated cells) and *Rgs13* (tuft cells), have been identified as risk genes for asthma in

[1]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. [2]Departments of Internal Medicine and Pediatrics, Pulmonary and Critical Care Unit, Massachusetts General Hospital, Boston, MA, USA. [3]Harvard Stem Cell Institute, Cambridge, MA, USA. [4]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. [6]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [7]Gregory Fleming James Cystic Fibrosis Research Center, Birmingham, AL, USA. [8]Department of Anatomy and Cell Biology, Carver College of Medicine, University of Iowa, Iowa City, IA, USA. [9]Department of Experimental Immunology, Academic Medical Center/University of Amsterdam, Amsterdam, The Netherlands. [10]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. [11]Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, MA, USA. [12]Department of Cell Biology, Duke University, Durham, NC, USA. [13]Duke Cancer Institute, Duke University, Durham, NC, USA. [14]Division of Pulmonary Critical Care, Department of Medicine, Duke University School of Medicine, Durham, NC, USA. [15]Regeneration Next, Duke University, Durham, NC, USA. [16]Department of Pediatrics, Massachusetts General Hospital, Boston, MA, USA. [17]Mucosal Immunology and Biology Research Center, Massachusetts General Hospital, Boston, MA, USA. [18]CFFT Lab, Cystic Fibrosis Foundation, Lexington, MA, USA. [19]Howard Hughes Medical Institute and Koch Institute for Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [20]These authors contributed equally: Daniel T. Montoro, Adam L. Haber, Moshe Biton. *e-mail: aregev@broadinstitute.org; jrajagopal@mgh.harvard.edu

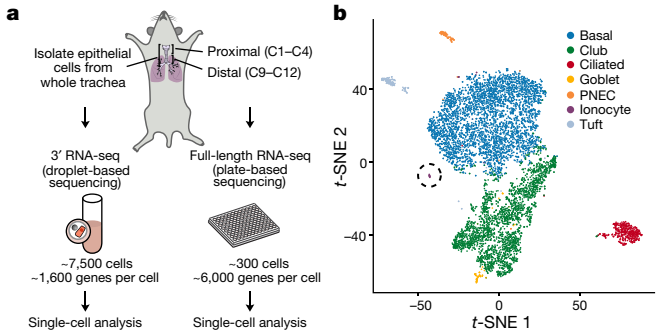


Fig. 1 | A single-cell expression atlas of mouse tracheal epithelial cells. **a**, Overview of the analysis. **b**, *t*-distributed stochastic neighbour embedding (*t*-SNE) of 7,193 3′ scRNA-seq profiles, coloured by cluster assignment and annotated post hoc. The ionocyte cluster is circled. PNEC, Pulmonary neuroendocrine cells.

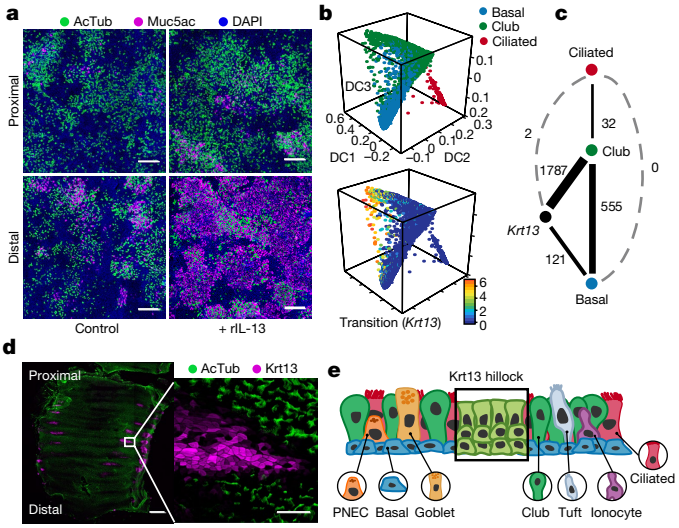


Fig. 2 | Club cell differentiation varies by location. **a**, Distal epithelia preferentially give rise to mucous metaplasia. Immunofluorescence showing cells positive for acetylated tubulin (AcTub; ciliated cells) and Muc5ac (goblet cells) in cultured epithelia from proximal (top panels) or distal (bottom panels) trachea stimulated with recombinant IL-13 (right) versus control (left). Scale bar, 200 μm. **b**, **c**, Differentiation trajectories. Diffusion map embedding (**b**) of 6,905 basal (blue), club (green) and ciliated (red) cells coloured by cluster assignment (top) or expression ($\log_2$(TPM+1), colour bar) of *Krt13* (bottom). **c**, Number of individual cells associated with each trajectory. **d**, **e**, Krt13$^+$ cells occur in hillock structures. **d**, Whole-mount stain of Krt13 (magenta) and acetylated tubulin (green), $n = 3$ mice. Scale bar: 500 μm (main), 50 μm (expanded inset). **e**, Schematic of squamous hillocks within pseudostratified ciliated epithelium.

genome-wide association studies (GWAS)[15] (Extended Data Fig. 3d–f). *Cdhr3* encodes a rhinovirus receptor and is associated with exacerbations in severe childhood asthma[16], suggesting that rhinovirus infection specifically of ciliated cells may precipitate exacerbations in some individuals. *Rgs13* is associated with asthma and IgE-mediated mast cell degranulation[17]; its specific expression in tuft cells implicates these cells as participants in asthmatic inflammation.

Mucous metaplasia (an excess of mucus-producing goblet cells) occurs more prominently in distal than in proximal mouse tracheal epithelium following allergen exposure[18]. Some cell-type-specific expression programs also vary along the proximodistal axis of the airway tree. Of 105 genes that are differentially expressed (FDR < 0.05, Mann–Whitney *U* test) between distal and proximal club cells (Extended Data Fig. 4a, Supplementary Table 5), distally enriched *Muc5b*[19], *Notch2*[20] and *Il13ra1*[21] are known to have roles in mucus metaplasia. Indeed, IL-13-induced mucous metaplasia in cultured epithelia resulted in greater goblet cell differentiation in distal epithelium (Fig. 2a, Extended Data Fig. 4b).

A cell population organized in ‘hillocks’

Cellular differentiation during homeostasis is an ongoing, asynchronous process. We inferred trajectories of cell differentiation using diffusion maps (Fig. 2b, c, Extended Data Fig. 4c), and characterized expression programs and transcription factors that vary coherently in transitional cells that were pseudotemporally ordered along trajectories that connect basal, club and ciliated cells (Extended Data Fig. 5, Supplementary Table 7, Methods). One of these trajectories reflects the known basal-to-club cell lineage path (DC1–DC2, $k = 555$ cells), but a second, distinct trajectory connects basal to club cells through a newly identified transitional cell (DC2–DC3, $k = 1{,}908$ cells) that uniquely expresses squamous epithelial markers *Krt4* and *Krt13* (FDR < 10^{-5}, LRT; Fig. 2b, c). The basal cell differentiation marker *Krt8*[22] does not distinguish the two paths that culminate in club cells (Extended Data Fig. 4c). We did not detect any cells transitioning from basal to ciliated cells (Fig. 2b, c), consistent with the homeostatic production of ciliated cells from club cells[1,22].

Surprisingly, many Krt13$^+$ cells are located in contiguous groups of stratified cells that lack luminal ciliated cells (Fig. 2d, e). Instead, luminal cells are Scgb1a1$^+$Krt13$^+$ club cells that lay atop Trp63$^+$Krt13$^+$ basal cells. Graded Trp63 expression extends from basal to suprabasal strata (Extended Data Fig. 4d). We term these unique structures ‘hillocks’. Labelling with 5-ethynyl-2′-deoxyuridine (EdU) was more concentrated in hillocks than in normal pseudostratified epithelium, indicating that hillocks are distinct zones of high turnover (Extended Data Fig. 4e, f). We generated *Scgb1a1-creER/LSL*-tdTomato mice to label all club cells, including hillock club cells. The fraction of labelled hillock club cells diminished with homeostatic turnover (Extended Data Fig. 4g), supporting a model in which Trp63$^+$Krt13$^+$ basal cells rapidly give rise to hillock club cells.

Hillock cells express regulators of cellular adhesion and squamous epithelial differentiation (*Ecm1*, *S100a11* and *Cldn3*), and genes associated with immunomodulation and asthma (*Lgals3* and *Anxa1*[23]; FDR < 10^{-10} LRT; Extended Data Fig. 4h, i, Supplementary Table 6). Overall, hillocks are characterized by rapid cellular turnover, squamous barrier function and immunomodulation.

Lineage tracing coupled to cellular dynamics

To monitor the generation of differentiated cell types, we developed ‘pulse-seq’, a novel assay that couples scRNA-seq and in vivo genetic lineage tracing over time (Extended Data Fig. 6a). We generated inducible *Krt5-creER/LSL*-mT/mG mice to label basal cells and their progeny with membrane-localized EGFP (mG), whereas non-lineage-labelled cells express membrane-localized tdTomato (mT). Following induction with tamoxifen, we profiled 66,265 mG$^+$ (Supplementary Fig. 1) and mT$^+$ cells by scRNA-seq at days 0, 30 and 60 of homeostatic turnover ($n = 9$ mice; three per time point). We identified the seven epithelial cell types and a population of proliferating cells, which were predominantly basal cells (Fig. 3a, Extended Data Fig. 6b). We calculated the fraction of lineage-labelled cells of each cell type at each time point (Fig. 3b, c) and estimated the daily labelling rate of each by quantile regression (Fig. 3d, Extended Data Fig. 6c, Methods). We then interpreted these data in the context of prior basal cell lineage traces in which club cells are labelled before ciliated cells[1,22], consistent with club cells being the direct parents of ciliated cells at homeostasis.

Initially, basal cells were specifically labelled (64.2%) with only infrequent labelling of rare cell types (<1.8%) and club cells (3.3%, $n = 3$ mice, Fig. 3b, c). Labelled club cells reflect a small population of transitional basal cells as they convert from a basal to club cell fate (Extended Data Fig 7e, f). The fraction of labelled cells among basal cells remains unchanged over time, consistent with self-renewal (Extended Data Fig. 6d). By contrast, the lineage-labelled fractions of tuft cells, neuroendocrine cells and ionocytes substantially increased (Fig. 3c), consistent with ongoing turnover. Rare-cell-type fractional lineage labelling

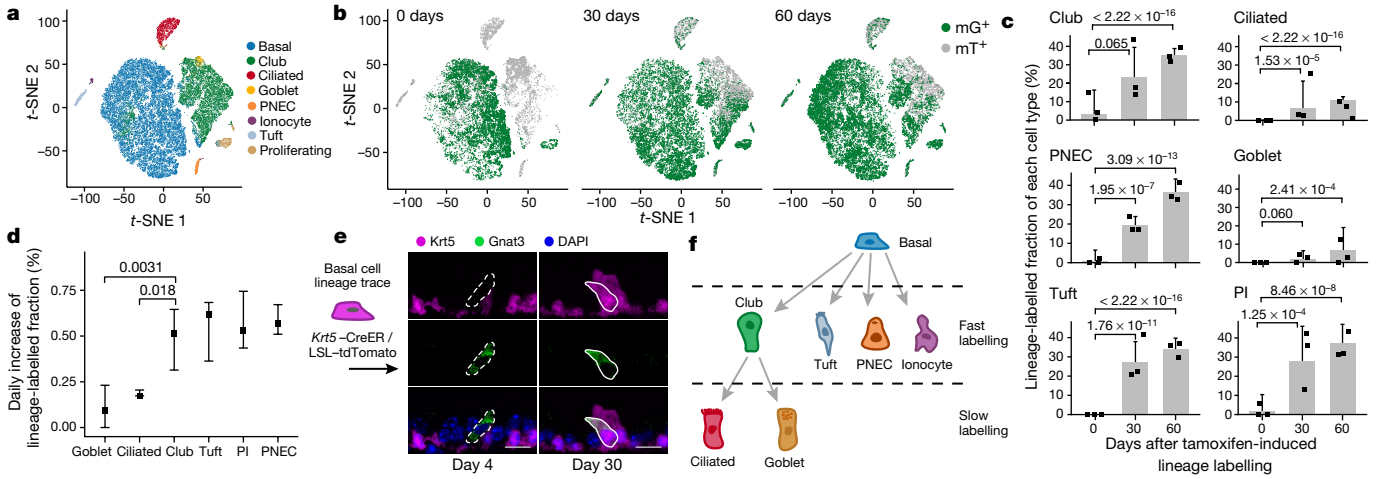


Fig. 3 | Tracking differentiation dynamics with pulse-seq. a, **b**, Pulse-seq tracks the sequential lineage labelling of each cell type. *t*-SNE visualization of 66,265 cells coloured by cluster assignment (**a**) or lineage label over time (**b**). mT, membrane tdTomato; mG, membrane EGFP. **c**, mG^+ lineage-labelled percentages of each tracheal epithelial cell type. Points represent the percentages of lineage-labelled cells for individual mice; bars show overall estimate; $n = 3$ mice per time point (*x* axis). Error bars, 95% CI; *P* values are indicated, LRT. PNEC, neuroendocrine; PI, pulmonary ionocyte. **d**, Estimated daily rate of lineage labelling (Extended Data Fig. 6c) by cell type. $n = 9$ mice; error bars, 95% CI; *P* values are indicated, rank test. **e**, Lineage validation in situ. Representative images of unlabelled (dashed outline) and basal lineage-labelled (solid outline) *Gnat3*$^+$ tuft cells. Scale bar, 10 μm. **f**, Cell types, lineage and cellular dynamics inferred using pulse-seq.

approximates that of club cells at day 30 and 60 (Fig. 3c, d), suggesting that these rare cell types, as with club cells, are immediate descendants of basal cells. This confirms a previous suggestion that solitary neuroendocrine cells are derived from basal cells[22]. By contrast, goblet and ciliated cells were labelled at a substantially lower rate (Fig. 3d), consistent with a model in which stem cells first produce club cells that, in turn, later produce goblet cells and ciliated cells.

We confirmed our lineage model with conventional in vivo lineage tracing with basal and club cell drivers. Over a 30-day basal cell lineage trace with *Krt5-creER*/*LSL*-tdTomato mice, the proportion of lineage-labelled tuft cells markedly increased (Fig. 3e, Extended Data Fig. 6e), whereas club cell lineage tracing with *Scgb1a1-creER*/*LSL*-tdTomato mice over the same time period labelled few of the tuft cells, and even fewer ionocytes or neuroendocrine cells, indicating that basal cells are the predominant source of these rare cell types (Fig. 3f, Extended Data Fig. 6f–h). We also investigated the turnover of the hillock club cells, identified by club cell sub-clustering (Extended Data Fig. 7a, b). The fraction of labelled hillock club cells grew more rapidly than the fraction of total labelled club cells (Extended Data Fig. 7c, d, g), consistent with the rapid turnover of hillocks.

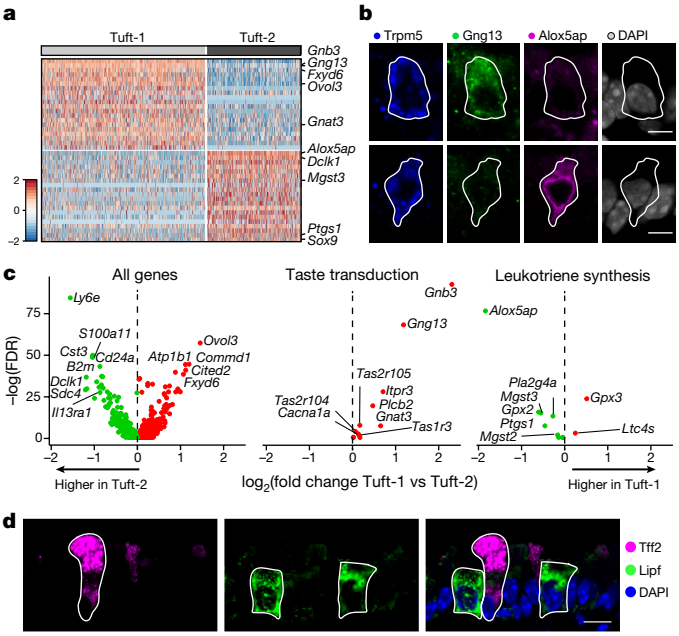


Fig. 4 | Tuft and goblet cell subtypes display unique functional gene expression programs. Tuft-1 and tuft-2 subclusters. **a**, Relative expression (row-wise *Z* score of $\log_2$(TPM+1)) of genes that are differentially expressed (FDR <0.25, LRT) between tuft cells of each subtype. **b**, Immunofluorescence of pan-tuft marker Trpm5 (blue) and tuft-1 (Gng13$^+$, green) or tuft-2 (Alox5ap$^+$, magenta) specific markers (cells are outlined) in vivo. DAPI, grey. $n = 3$ mice, four replicate trachea sections were examined for each mouse. Scale bar, 5 μm. **c**, Distinct expression programs in tuft-1 and tuft-2 cells. Differential expression in tuft cell subtypes for all genes (left), taste genes (centre) and leukotriene synthesis genes (right). Labelled genes are differentially expressed (FDR <0.01, LRT); $k = 892$ cells; $n = 15$ mice. **d**, Immunofluorescence validation of goblet-1 (Tff2, magenta) and goblet-2 (Lipf, green) cells (solid outlines). DAPI, blue; $n = 3$ mice, four replicate trachea sections were examined for each mouse. Scale bar, 10 μm.

Distinct subsets of tuft and goblet cells

Tuft cells express the largest number of specific G-protein-coupled receptors and taste receptors, consistent with a sensory function (FDR < 0.001, LRT; Extended Data Fig. 8a, b, Supplementary Table 4). Airway tuft cells express the alarmins *Il25* and *Tslp* (FDR < 10^{-3}, Extended Data Fig. 8c), which initiate type-2 immunity in the gut[24], and possess lateral cytoplasmic extensions (Extended Data Fig. 8d) that may extend their chemosensory span.

Next, we separately re-clustered each rare cell type after aggregating both droplet-based datasets (Figs. 1b, 3a, $n = 15$ mice). Tuft cells partitioned into three clusters: immature tuft, tuft-1 and tuft-2 cells (Fig. 4a, Extended Data Fig. 8e, f, i, Supplementary Table 8). Tuft-1 cells expressed genes associated with taste transduction ($P = 2.07 \times 10^{-14}$, hypergeometric test), whereas tuft-2 cells expressed genes that mediate leukotriene biosynthesis, notably *Alox5ap*[25] ($P = 3.13 \times 10^{-4}$, hypergeometric test), which are central mediators of inflammation and asthma (Fig. 4a–c). As in the gut[24], tuft-2 cells are also enriched for immune cell-associated *Ptprc* (CD45, FDR = 0.064, LRT). Both tuft cell subsets are generated at similar rates by basal cells (Extended Data Fig. 8g), but canonical tuft cell transcription factors are associated with specific subsets of tuft cells, including *Pou2f3* (tuft-1) and *Gfi1b*, *Spib*, and *Sox9* (tuft-2, FDR <0.01, LRT, Extended Data Fig. 8h).

The most highly enriched marker across goblet cells was *Gp2* (Extended Data Fig. 1e, Supplementary Table 1), a marker of intestinal M cells associated with mucosal immunity[26]. Goblet cells

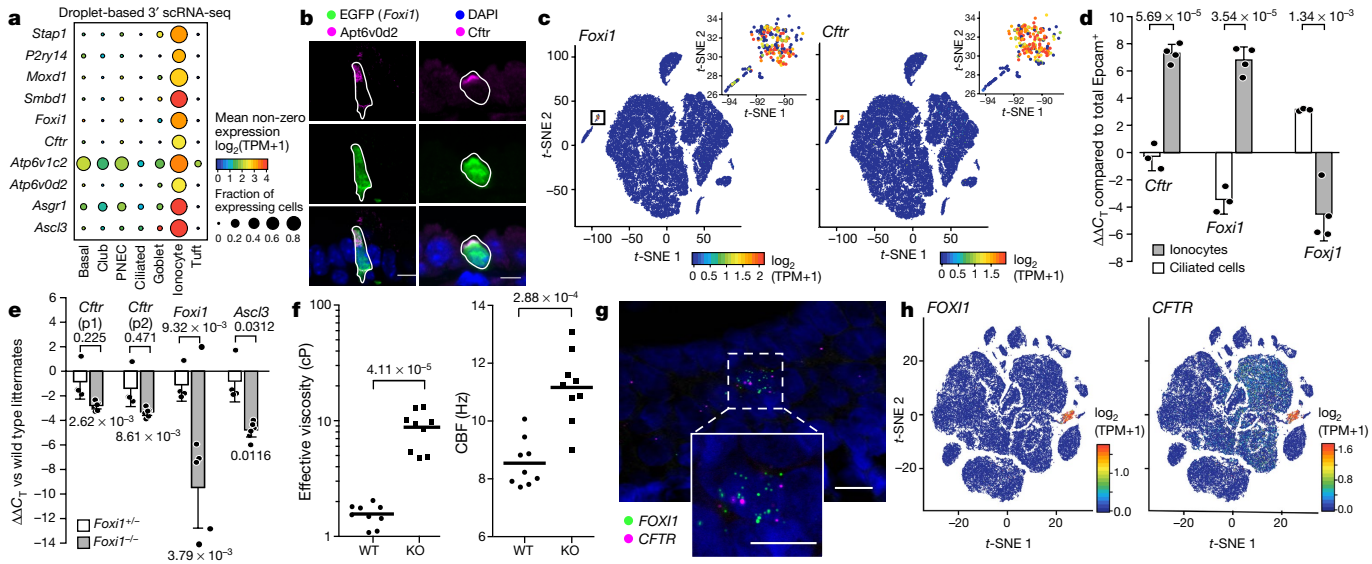


Fig. 5 | The pulmonary ionocyte is a novel mouse and human airway epithelial cell type that specifically expresses CFTR. a, Mouse pulmonary ionocyte markers. Expression level of ionocyte markers (FDR < 0.05, LRT, 3′ scRNA-seq dataset) in each airway epithelial cell type. *Smbd1* was formerly known as *Gm933*. **b**, Immunofluorescence co-labelling of EGFP $(Foxi1)^+$ ionocytes (solid outline), Atp6v0d2 (left) and Cftr (right). DAPI, blue; $n = 3$ mice, four replicate trachea sections were examined for each mouse. Scale bar: 10 μm (left), 5 μm (right). **c**, *t*-SNE plot of 66,265 pulse-seq cells and ionocyte subset (black box, inset) coloured by expression of ionocyte markers *Foxi1* (left) and *Cftr* (right). **d**, qRT–PCR confirms ionocyte enrichment of *Cftr*. Expression ($\Delta\Delta C_T$, Supplementary Table 12) of ionocyte (*Cftr*, *Foxi1*) and ciliated cell (*Foxj1*) markers (*x* axis) in ionocytes and ciliated cells isolated from *Foxi1*-EGFP ($n = 4$, dots) and *Foxj1*-EGFP mice ($n = 3$), respectively. Samples are normalized to $EpCAM^+$ populations from wild-type mice ($n = 6$). Error bars, 95% CI; *t*-test, two-sided. *P* values are indicated. **e**, *Foxi1* knockout decreases expression of ionocyte transcription factors and *Cftr* in air–liquid interface (ALI)-cultured epithelia. Expression ($\Delta\Delta C_T$, Supplementary Table 12) of ionocyte markers in heterozygous ($Foxi1^{+/-}$, $n = 4$) and homozygous knockouts ($Foxi1^{-/-}$, $n = 6$), normalized to wild-type littermates ($n = 8$). Error bars, 95% CI; *P* values are indicated, Holm–Sidak test. **f**, *Foxi1* knockout disrupts mucosal homeostasis in ALI-cultured epithelia. Effective viscosity (left) and ciliary beat frequency (right) assayed with μOCT in homozygous *Foxi1*(KO) ($n = 9$, dots) versus wild-type (WT) littermates ($n = 9$ mice). Bars show mean. *P* values are indicated, Mann–Whitney *U* test. **g**, **h**, Human pulmonary ionocytes are the main source of *CFTR* in human bronchial epithelium. Human ionocytes detected by fluorescent in situ hybridization of *FOXI1* and *CFTR* in bronchi (**g**; $n = 3$ bronchi). *t*-SNE of 78,217 3′ droplet scRNA-seq profiles (points) from human bronchial epithelium (**h**; $n = 1$ patient), coloured by expression of *FOXI1* (left) and *CFTR* (right). Scale bar, 10 μm.

partitioned into three subsets, immature goblet, goblet-1 and goblet-2 (Extended Data Fig. 8j–l, Supplementary Table 8). Goblet-1 cells are enriched for the expression of genes encoding key mucosal proteins (*Tff1*, *Tff2* and *Muc5b*[19], FDR <0.001, LRT) and secretory regulators (for example, *Lman1l* or *P2rx4*[27], FDR <0.1, LRT). We confirmed the co-expression of *Tff2* and *Muc5ac* in goblet-1 cells by antibody staining (Extended Data Fig. 8m). Goblet-2 cells specifically express *Dcpp1*, *Dcpp2* and *Dcpp3*, orthologues of *ZG16B*, which codes for a lectin-like secreted protein that aggregates bacteria[28], and *Lipf*, a secreted gastric lipase that hydrolyses triglycerides. We identified unique $Tff2^+$ goblet-1 and $Lipf^+$ goblet-2 cells by immunostaining (Fig. 4d).

Foxi1$^+$ pulmonary ionocytes highly express Cftr

We confirmed that ionocytes are a newly identified cell population in vivo using transgenic *Foxi1*-EGFP reporter mice and Foxi1 immunoreactivity. EGFP (*Foxi1*) co-localizes with global airway markers (Sox2 and Tff1), but not markers of the other cell types (Extended Data Fig. 9a). On average, we detected 1,038 ± 501 ionocytes in the surface epithelium of each mouse trachea ($n = 3$ mice, Extended Data Fig. 9b), accounting for <1% of epithelial cells.

Pulmonary ionocytes specifically express the V-ATPase-subunit genes *Atp6v1c2* and *Atp6v0d2* (FDR < 0.0005, LRT, Fig. 5a, Extended Data Figs. 3b, 9c, Supplementary Table 1) and are uniquely immunoreactive for ATP6v0d2 (Fig. 5b). This profile resembles that of *Xenopus* and zebrafish skin ionocytes, in which *Foxi1* orthologues specify cell identity and regulate V-ATPase expression[8,9]. Mouse Foxi1 also controls the expression of V-ATPase—which is important for ion transport and fluid pH[29]—in specialized cells of the inner ear, kidney and epididymis. Like zebrafish ionocytes[30], pulmonary ionocytes extend lateral processes (Extended Data Fig. 9d) that may be involved in chemosensation or cell-to-cell communication.

Pulmonary ionocytes specifically express the cystic fibrosis transmembrane conductance regulator (*Cftr*) gene (FDR = 0.00103, initial droplet data; FDR = 0.000361; pulse-seq, LRT, Fig. 5a, c, Extended Data Figs. 3b, 9c, Supplementary Tables 1–3). Ionocytes comprise only 0.42% of the mouse cells profiled by scRNA-seq, but express 54.4% of all detected *Cftr* transcripts. For comparison, the vastly more abundant ciliated cells express 1.5% of total *Cftr* transcripts. Additionally, EGFP $(Foxi1)^+$ ionocytes were specifically labelled by Cftr antibody (Fig. 5b). We further confirmed ionocyte-specific enrichment of *Cftr* by quantitative PCR with reverse transcription (qRT–PCR) analysis of the mRNA of prospectively isolated populations of primary ionocytes and ciliated cells (191.6-fold enrichment) or bulk $EpCAM^+$ epithelial cells (158.1-fold enrichment, Fig. 5d, Supplementary Table 12).

We detected ionocytes in mouse submucosal glands, structures associated with cystic fibrosis pathogenesis[31,32], and in nasal and olfactory epithelia (Extended Data Fig. 9e–g). Ionocytes specifically express cochlin (Supplementary Table 1), a secreted protein that confers antibacterial activity against the two most prominent pathogens in cystic fibrosis lung disease[33]. Using *Foxi1* knockout (*Foxi1*(KO)) mouse epithelia, we show that *Foxi1* is required for expression of the ionocyte transcription factor *Ascl3* (96.3% reduction) and the majority of *Cftr* expression (87.6% reduction, Fig. 5e, Supplementary Table 12). Mouse epithelia deficient in *Ascl3* display moderately reduced *Foxi1* and *Cftr* expression (Extended Data Fig. 10a).

Ionocytes regulate airway surface physiology

Tight control of airway surface liquid (ASL) and mucus viscosity is necessary for effective mucociliary clearance and is disrupted in cystic

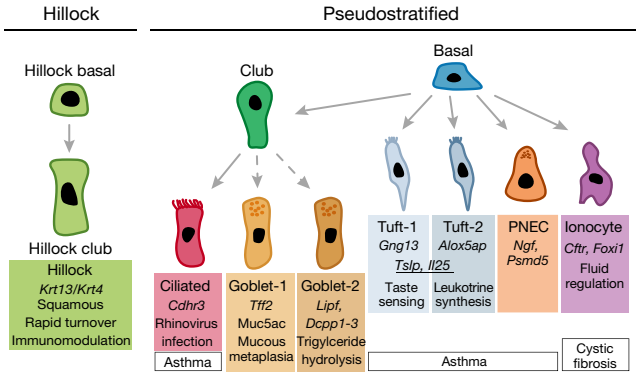


Fig. 6 | Lineage hierarchy of the airway epithelium. Specific cells are associated with novel cell-type markers, pathways and diseases.

fibrosis[34,35]. We assessed ASL height, mucus viscosity and ciliary beat frequency in polarized *Foxi1*(KO) mouse airway epithelia using live imaging by micro-optical coherence tomography (μOCT) and particle-tracking microrheology (Methods). We found increased reflectance intensity (Extended Data Fig. 10b) and increased effective viscosity of airway mucus (Fig. 5f) in *Foxi1*(KO) mice, consistent with animal models of cystic fibrosis[34,36]. Ciliary beat frequency also increased in the *Foxi1*(KO) epithelium (Fig. 5f), consistent with a response to an elevated mechanical load due to the increased mucus viscosity[37]. As with some mouse *Cftr*(KO) models[38,39], neither depth nor pH (Extended Data Fig. 10c, d) of the ASL was significantly altered in *Foxi1*(KO) epithelial cultures.

We also tested whether *Foxi1*(KO) epithelia produce abnormal forskolin-induced and CFTR inhibitor ($CFTR_{inh}$-172)-blocked equivalent currents (ΔI_{eq}) in Ussing chambers (Methods). *Foxi1*(KO) mouse epithelium lacks *Cftr* (Fig. 5e), yet displayed increases in $CFTR_{inh}$-172-inhibitable forskolin currents (Extended Data Fig. 10e, f), similar to the compensatory currents observed in *Cftr*-mutant mice[40].

We further investigated the role of *Foxi1* in ferrets, a species that models cystic fibrosis well[41]. CRISPR–dCas9–VP64-p65-mediated transcriptional activation of *Foxi1* (*Foxi1*(TA)) increased airway epithelial expression of *Cftr* and other ionocyte genes (Extended Data Fig. 10g). *Foxi1*(TA) cultures displayed increased forskolin-induced short-circuit currents (ΔI_{sc}) and CFTR inhibitor (GlyH 101)-induced ΔI_{sc} relative to mock-transfected controls (Extended Data Fig. 10h, i). Therefore, Foxi1 regulates CFTR expression and function in ferret airway epithelium.

Human pulmonary ionocytes are *CFTR*-rich

Human pulmonary ionocytes are the major source of *CFTR* expression in the airway epithelium. We detected rare $FOXI1^{+}CFTR^{+}$ cells in human bronchi using RNA fluorescent in situ hybridization (Fig. 5g) Additionally, we detected 765 ionocytes by unsupervised clustering of 87,285 primary human airway cells analysed by scRNA-seq (unpublished data). Human ionocytes comprise 0.5–1.5% of epithelial cells along the conducting airways (Supplementary Table 10) and specifically express *FOXI1*, *ASCL3* and *CFTR* (FDR $< 10^{-10}$, LRT; Fig. 5h, Extended Data Fig. 10j, Supplementary Table 11), whereas scattered basal and secretory cells express low levels of *CFTR*. *FOXI1* transcriptional activation increases ionocyte-specific gene expression in human airway epithelial cultures[7].

Discussion

Our single-cell atlas of mouse tracheal epithelium identified: (1) a new cell type, the ionocyte; (2) new subclasses of disease-relevant tuft and goblet cells; and (3) novel transitional cells arranged in discrete high turnover structures that we named 'hillocks' (Fig. 6). Our pulse-seq analysis further illuminated the differentiation dynamics of this new hierarchy of cells. The analysis revealed a simple model of epithelial turnover in which solitary neuroendocrine cells, tuft cells, ionocytes and club cells are all produced at the same rate by basal cells. We speculate that the high turnover hillocks represent injury-responsive structures that couple immunomodulation and barrier function.

The pulmonary ionocyte bears the hallmarks of an ancient prototype cell. The ionocyte occurs in animals as distinct as fish, frog and human, and is associated with a particular physiologic function: fluid regulation at the epithelial interface. We show that $Foxi1^{+}Cftr^{+}$ ionocytes reside at multiple levels of the airway tree and that they are responsible for the majority of *Cftr* expression. Indeed, we demonstrate that they need to function correctly to maintain airway surface physiology, including mucus viscosity.

Increased forskolin-inducible currents in *Foxi1*(KO) mice are consistent with the compensatory activation of forskolin-inducible currents in large airway epithelia of *Cftr*-mutant mice[40]. These currents may moderate the severity of the mouse cystic fibrosis phenotype, and the channels responsible could serve as therapeutic targets.

Since human pulmonary ionocytes express higher levels of *CFTR* than any other large airway cell type, the current understanding of the cellular basis of cystic fibrosis is likely to be incomplete. Previous studies have shown that whereas the nasal epithelia of *Cftr*-null mice phenocopy ion transport abnormalities of human cystic fibrosis airways[42], expression of CFTR in ciliated cells does not rescue these abnormalities[43]. Taken together with our findings, this suggests that the correct cellular context of CFTR expression may be required for proper CFTR function. As we show that existing ionocytes are replaced by new ionocytes generated from basal stem cells, we speculate that these basal cells are the appropriate long-lasting cellular targets for cystic fibrosis gene therapy. Studies of single-cell expression patterns in cells from human patients with cystic fibrosis will help further address these questions.

In sum, we present a new cellular narrative of airways disease, in which specific cell types and subtypes are associated with particular disease genes. Because lineage paths and cell states may be substantially altered in disease states, comprehensive cell atlases of both healthy and diseased human lung are needed[44] as a prelude to reframing the biology and pathobiology of the lung and its diseases.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0393-7

Received: 31 July 2017; Accepted: 21 June 2018;
Published online 1 August 2018.

1. Rock, J. R. et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
2. Lump, A. *Nunn's Applied Respiratory Physiology*, 8th edn (Elsevier, Edinburgh, 2016)
3. Ardini-Poleske, M. E. et al. LungMAP: The Molecular Atlas of Lung Development Program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **313**, L733–L740 (2017).
4. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
5. Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118–1123 (2018).
6. Zepp, J. A. et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* **170**, 1134–1148 (2017).
7. Plasschaert, L. W. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* https://doi.org/10.1038/s41586-018-0394-6 (2018).
8. Quigley, I. K., Stubbs, J. L. & Kintner, C. Specification of ion transport cells in the *Xenopus* larval skin. *Development* **138**, 705–714 (2011).
9. Esaki, M. et al. Mechanism of development of ionocytes rich in vacuolar-type H^{+}-ATPase in the skin of zebrafish larvae. *Dev. Biol.* **329**, 116–129 (2009).
10. Pardo-Saganta, A. et al. Parent stem cells can serve as niches for their daughter cells. *Nature* **523**, 597–601 (2015).
11. Tsao, P.-N. et al. Epithelial Notch signaling regulates lung alveolar morphogenesis and airway epithelial integrity. *Proc. Natl Acad. Sci. USA* **113**, 8242–8247 (2016).
12. Sriuranpong, V. et al. Notch signaling induces rapid degradation of achaete-scute homolog 1. *Mol. Cell. Biol.* **22**, 3129–3139 (2002).

13. Moriyama, M. et al. Multiple roles of Notch signaling in the regulation of epidermal development. *Dev. Cell* **14**, 594–604 (2008).
14. Verzi, M. P., Khan, A. H., Ito, S. & Shivdasani, R. A. Transcription factor Foxq1 controls mucin gene expression and granule content in mouse stomach surface mucous cells. *Gastroenterology* **135**, 591–600 (2008).
15. Li, M. J. et al. GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* **44**, D869–D876 (2016).
16. Bochkov, Y. A. et al. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proc. Natl Acad. Sci. USA* **112**, 5485–5490 (2015).
17. Bansal, G., Xie, Z., Rao, S., Nocka, K. H. & Druey, K. M. Suppression of immunoglobulin E-mediated allergic responses by regulator of G protein signaling 13. *Nat. Immunol.* **9**, 73–80 (2008).
18. Pardo-Saganta, A., Law, B. M., Gonzalez-Celeiro, M., Vinarsky, V. & Rajagopal, J. Ciliated cells of pseudostratified airway epithelium do not become mucous cells after ovalbumin challenge. *Am. J. Respir. Cell Mol. Biol.* **48**, 364–373 (2013).
19. Roy, M. G. et al. Muc5b is required for airway defence. *Nature* **505**, 412–416 (2014).
20. Danahay, H. et al. Notch2 is required for inflammatory cytokine-driven goblet cell metaplasia in the lung. *Cell Reports* **10**, 239–252 (2015).
21. Munitz, A., Brandt, E. B., Mingler, M., Finkelman, F. D. & Rothenberg, M. E. Distinct roles for IL-13 and IL-4 via IL-13 receptor $\alpha1$ and the type II IL-4 receptor in asthma pathogenesis. *Proc. Natl Acad. Sci. USA* **105**, 7240–7245 (2008).
22. Watson, J. K. et al. Clonal dynamics reveal two distinct populations of basal cells in slow-turnover airway epithelium. *Cell Reports* **12**, 90–101 (2015).
23. Ng, F. S. P. et al. Annexin-1-deficient mice exhibit spontaneous airway hyperresponsiveness and exacerbated allergen-specific antibody responses in a mouse model of asthma. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* **41**, 1793–1803 (2011).
24. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
25. Dixon, R. A. et al. Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. *Nature* **343**, 282–284 (1990).
26. Hase, K. et al. Uptake through glycoprotein 2 of $FimH^+$ bacteria by M cells initiates mucosal immune response. *Nature* **462**, 226–230 (2009).
27. Miklavc, P., Thompson, K. E. & Frick, M. A new role for $P2X_4$ receptors as modulators of lung surfactant secretion. *Front. Cell. Neurosci.* **7**, 171 (2013).
28. Bergström, J. H. et al. Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. *Proc. Natl Acad. Sci. USA* **113**, 13833–13838 (2016).
29. Vidarsson, H. et al. The forkhead transcription factor Foxi1 is a master regulator of vacuolar H-ATPase proton pump subunits in the inner ear, kidney and epididymis. *PLoS ONE* **4**, e4471 (2009).
30. Jonz, M. G. & Nurse, C. A. Epithelial mitochondria-rich cells and associated innervation in adult and developing zebrafish. *J. Comp. Neurol.* **497**, 817–832 (2006).
31. Hoegger, M. J. et al. Impaired mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis. *Science* **345**, 818–822 (2014).
32. Engelhardt, J. F. et al. Submucosal glands are the predominant site of CFTR expression in the human bronchus. *Nat. Genet.* **2**, 240–248 (1992).
33. Py, B. F. et al. Cochlin produced by follicular dendritic cells promotes antibacterial innate immunity. *Immunity* **38**, 1063–1072 (2013).
34. Birket, S. E. et al. A functional anatomic defect of the cystic fibrosis airway. *Am. J. Respir. Crit. Care Med.* **190**, 421–432 (2014).
35. Birket, S. E. et al. Development of an airway mucus defect in the cystic fibrosis rat. *JCI Insight* **3**, e97199 (2018).
36. Tang, X. X. et al. Acidic pH increases airway surface liquid viscosity in cystic fibrosis. *J. Clin. Invest.* **126**, 879–891 (2016).
37. Liu, L. et al. An autoregulatory mechanism governing mucociliary transport is sensitive to mucus load. *Am. J. Respir. Cell Mol. Biol.* **51**, 485–493 (2014).
38. Shah, V. S. et al. Airway acidification initiates host defense abnormalities in cystic fibrosis mice. *Science* **351**, 503–507 (2016).
39. Tarran, R. et al. Regulation of murine airway surface liquid volume by CFTR and Ca^{2+}-activated Cl^- conductances. *J. Gen. Physiol.* **120**, 407–418 (2002).
40. Liu, X., Yan, Z., Luo, M. & Engelhardt, J. F. Species-specific differences in mouse and human airway epithelial biology of recombinant adeno-associated virus transduction. *Am. J. Respir. Cell Mol. Biol.* **34**, 56–64 (2006).
41. Sun, X. et al. Disease phenotype of a ferret CFTR-knockout model of cystic fibrosis. *J. Clin. Invest.* **120**, 3149–3160 (2010).
42. McCarron, A., Donnelley, M. & Parsons, D. Airway disease phenotypes in animal models of cystic fibrosis. *Respir. Res.* **19**, 54 (2018).
43. Ostrowski, L. E. et al. Expression of CFTR from a ciliated cell-specific promoter is ineffective at correcting nasal potential difference in CF mice. *Gene Ther.* **14**, 1492–1501 (2007).
44. Regev, A. et al. Science Forum: The Human Cell Atlas. *eLife* **6**, e27041 (2017).

Acknowledgements We thank L. Gaffney for help with figure preparation, P. Oyem for assistance with electrophysiological assays, the New England Organ Bank for facilitating the acquisition of donor lungs, and the HSCI Flow Cytometry Core and CNY Flow Cytometry Core facilities. This work was supported by the Klarman Cell Observatory at the Broad Institute (A.R. and J.R.), the Manton Foundation (A.R.), HHMI (A.R. and J.R.), New York Stem Cell Foundation (J.R.), NIH-NHLBI (J.R.), the Ludwig Cancer Institute at Harvard (J.R.), and the Harvard Stem Cell Institute (J.R.). M.B. was supported by a postdoctoral fellowship from the Human Frontiers Science Program. D.T.M. was supported by a predoctoral fellowship from NIH-NHLBI 1F31HL136128-01. P.R.T. is a Whitehead Scholar and was supported by a career development award from NHLBI/NIH (4R00HL127181) and funds from Regeneration NeXT Initiative at Duke University. S.M.R. was supported by NIH P30 DK072482 and R35 HL135816. J.F.E. was supported by P01 HL051670, R24 HL123482 and R01 DK047967. J.R. is a MGH Maroni Research Scholar, a Harrington Investigator of the NYSCF and HHMI Faculty Scholar.

Reviewer information *Nature* thanks I. Amit and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.T.M., A.L.H., M.B., A.R. and J.R. conceived the study; J.R. and A.R. supervised research; A.L.H. designed and performed computational analysis; D.T.M. designed, carried out and analysed experiments with V.V., B.L., S.C., J.V. and P.R.T.; M.B. advised on experimental design and performed mouse single-cell experiments with N.R., G.B., L.N. and D.D.; H.B. and M.M. provided mouse electrophysiology data; S.B., H.M.L., G.J.T. and S.M.R. performed and interpreted μOCT experiments; S.B. performed and interpreted pH experiments. F.Y. and J.F.E. performed and interpreted ferret expression and electrophysiology data; A.T., A.W., M.SI., J.W. and O.R.-R. contributed human single-cell data and analysis; H.M. assisted with cell culture. M.Sh. previously observed $Krt13^+$ cells arranged as hillocks. D.T.M., A.L.H., A.R. and J.R. wrote the manuscript with input from all authors.

Competing interests A.R. is a member of the SAB for Thermo Fisher Scientific, Syros Pharmaceuticals and Driver Group, and a founder of Celsius Therapeutics. D.T.M., A.L.H., M.B., O.R.-R., A.R. and J.R. are co-inventors on PCT/US2018/027337 filed by the Broad Institute relating to innovative advances in epithelial hierarchy and ionocytes described in this manuscript.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0393-7.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0393-7.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.R. or J.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mouse models. The MGH Subcommittee on Research Animal Care approved animal protocols in accordance with NIH guidelines. *Krt5-creER*[1] and *Scgb1a1-creER*[45] mice have previously been described. *Foxi1*-EGFP mice were purchased from GENSAT. C57BL/6J mice (stock no. 000664), LSL-mT/mG mice (mouse stock no. 007676), and LSL-tdTomato (stock no. 007914), Ascl3-EGFP-Cre mice (stock no. 021794) and Foxi1(KO) mice (stock no. 024173) were purchased from the Jackson Laboratory. To label basal cells and secretory cells for in vivo lineage traces, we administered tamoxifen by intraperitoneal injection (3 mg per 20 g body weight) three times every 48 h to induce the Cre-mediated excision of a stop codon and subsequent expression of tdTomato. For pulse-seq experiments, we administered tamoxifen by intraperitoneal injection (2 mg per 20 g body weight) three times every 24 h to induce the Cre-mediated excision of a stop codon and subsequent expression of EGFP. To label proliferating cells, we administered 5-ethynyl-2′-deoxyuridine (EdU) per 25g mouse by intraperitoneal injection (2 mg per 20 g body weight). Six-to-twelve-week-old mice were used for all experiments. Male C57BL/6 mice were used for the full length and initial 3′ scRNA-seq experiments. Both male and female mice were used for lineage tracing and pulse-seq experiments. We used three mice for each lineage time point.

Littermate mice were assigned into groups on the basis of genotype.

Immunofluorescence, microscopy and cell counting. Tracheae were dissected and fixed in 4% PFA for 2 h at 4 °C followed by two washes in PBS, and then embedded in μOCT. Cryosections (6 μm) were treated for epitope retrieval with 10 mM citrate buffer at 95 °C for 10–15 min, permeabilized with 0.1% Triton X-100 in PBS, blocked in 1% BSA for 30 min at room temperature (27 °C), incubated with primary antibodies for 1 h at room temperature, washed, incubated with appropriate secondary antibodies diluted in blocking buffer for 1 h at room temperature, washed and counterstained with DAPI.

In the case of whole-mount trachea stains, tracheas were longitudinally resectioned along the posterior membrane, permeabilized with 0.3% Triton X-100 in PBS, blocked in 0.3% BSA and 0.3% Triton X-100 for 120 min at 37 °C on an orbital shaker, incubated with primary antibodies for 12 h at 37 °C (again on an orbital shaker), washed in 0.3% Triton X-100 in PBS, incubated with appropriate secondary antibodies diluted in blocking buffer for 1 h at 37 °C, washed in 0.3% Triton X-100 in PBS and counterstained with Hoechst 33342. They were then mounted on a slide between two magnets to ensure a flat imaging surface.

The following antibodies were used: rabbit anti-Atp6v0d2 (1/300; pa5-44359, Thermo), goat anti-CC10 (aka Scgb1a1, 1:500; SC-9772, Santa Cruz), anti-mouse CD45-PE (1/500; #12-0451-83, eBioscience), hamster anti-CD81(1/500; MA1-70091, Thermo), rabbit anti-CFTR (1:100; ACL-006, Alomone), mouse anti-Chromogranin A (1/500; sc-393941, Santa Cruz), rat anti-Cochlin (1/500; MABF267, Millipore), anti-mouse EpCAM-PECy7 (1/500; 324221, Biolegend), goat anti-FLAP (aka Alox5ap, 1:500; NB300-891, Novus), goat anti-Foxi1 (1:250; ab20454, Abcam), chicken anti-GFP (1:500; GFP-1020, Aves Labs), rabbit anti-Gnat3 (1/300; sc-395, Santa Cruz), rabbit anti-Gng13 (1:500; ab126562, Abcam), rabbit anti-Krt13 (1/500; ab92551, Abcam), goat anti-Krt13 (1/500; ab79279, Abcam), goat anti-Lipf (1:100; MBS421137, mybiosource.com), mouse anti-Muc5ac (1/500; ma1-38223, Thermo), mouse anti-Muc5ac (1/500; ma1-38223, Thermo), mouse anti-p63 (1:250; gtx102425, GeneTex), rabbit anti-Tff2 (1/500; 13681-1-AP, ProteinTech), rabbit anti-Trpm5 (1:500; ACC-045, Alamone), mouse anti-tubulin, acetylated (1:100; T6793, Sigma). All secondary antibodies were Alexa Fluor conjugates (488, 594 and 647) and used at 1:500 dilution (Life Technologies): dk anti-chicken 488 A-11039, dk anti-goat 488 A-11055, dk anti-mouse 488 A-21202, dk anti-rabbit 488 A-21206, dk anti-rat 488 A-21208, dk anti-goat 594 A-11058, dk anti-mouse 594 R37115, dk anti-rabbit 594 R37119, dk anti-hamster 647 A-21451, dk anti-goat 647 A-21447, dk anti-mouse 647 A-31571, dk anti-rabbit 647 A-31573.

EdU was stained in fixed sections alongside the above antibody stains as previously described[46].

Confocal images for both slides and whole-mount tracheas were obtained with an Olympus FV10i confocal laser-scanning microscope with a 60× oil objective. Cells were manually counted based on immunofluorescence staining of markers for each of the respective cell types. Cartilage rings (1 to 12) were used as reference points in all the tracheal samples to count specific cell types on the basis of immunostaining. Serial sections were stained for the antibodies tested and randomly selected slides were used for cell counting.

Cell dissociation and FACS. Airway epithelial cells from trachea were dissociated using papain solution. For whole-trachea sorting, longitudinal halves of the trachea were cut into five pieces and incubated in papain dissociation solution and incubated at 37 °C for 2 h. For proximal–distal cell sorting, proximal (cartilage 1–4) and distal (cartilage 9–12) trachea regions were dissected and dissociated by papain independently. After incubation, dissociated tissues were passed through a cell strainer and centrifuged and pelleted at 500*g* for 5 min. Cell pellets were dispersed and incubated with Ovo-mucoid protease inhibitor (Worthington Biochemical, cat. no. LK003182) to inactivate residual papain activity by incubating on a rocker at 4 °C for 20 min. Cells were then pelleted and stained with EpCAM–PECy7 (1:50; 25-5791-80, eBioscience) and CD45, CD81, or on the basis of EGFP expression for 30 min in 2.5% FBS in PBS on ice. After washing, cells were sorted by fluorescence (antibody staining and/or EGFP) on a BD FACS Aria (BD Biosciences) using FACS Diva software and analysis was performed using FlowJo (version 10) software.

For plate-based scRNA-seq, single cells were sorted into each well of a 96-well PCR plate containing 5 μl of TCL buffer with 1% 2-mercaptoenthanol. In addition, a population control of 200 cells was sorted into one well and a no-cell control was sorted into another well. After sorting, the plate was sealed with a Microseal F, centrifuged at 800*g* for 1 min and immediately frozen on dry ice. Plates were stored at −80 °C until lysate cleanup.

For droplet-based scRNA-seq, cells were sorted into an Eppendorf tube containing 50 μl of 0.4% BSA in PBS and stored on ice until proceeding to the GemCode Single Cell Platform.

Plate-based scRNA-seq. Single cells were processed using a modified SMART-seq2 protocol as previously described[47]. In brief, RNAClean XP beads (Agencourt) were used for RNA lysate cleanup, followed by reverse transcription using Maxima Reverse Transcriptase (Life Technologies), whole transcription amplification (WTA) with KAPA HotStart HIFI 2X ReadyMix (Kapa Biosystems) for 21 cycles and purification using AMPure XP beads (Agencourt). WTA products were quantified with Qubit dsDNA HS Assay Kit (Thermo Fisher), visualized with high sensitivity DNA Analysis Kit (Agilent) and libraries were constructed using Nextera XT DNA Library Preparation Kit (Illumina). Population and no-cell controls were processed with the same methods as single cells. Libraries were sequenced on an Illumina NextSeq 500.

Droplet-based scRNA-seq. Single cells were processed through the GemCode Single Cell Platform using the GemCode Gel Bead, Chip and Library Kits (V1) or single-cell suspensions were loaded onto 3′ library chips for the Chromium Single Cell 3′ Library (V2, PN-120233) according to the manufacturer's recommendations (10X Genomics). In brief, single cells were partitioned into Gel Beads in Emulsion in the GemCode instrument with cell lysis and barcoded reverse transcription of RNA, followed by amplification, shearing and 5′ adaptor and sample index attachment. An input of 6,000 single cells was added to each channel with a recovery rate of roughly 1,500 cells. Libraries were sequenced on an Illumina Nextseq 500.

qRT–PCR. Cells isolated by FACS were sorted into 150 μl TRIzol LS (Thermo Fisher Scientific), whereas ALI culture membranes were submerged in 300 μl of standard TRIzol solution (Thermo Fisher Scientific). A standard chloroform extraction was performed, followed by an RNeasy column-based RNA purification (Qiagen) according to the manufacturer's instructions. When possible, 1 μg (otherwise 100 ng) RNA was converted to cDNA using SuperScript VILO kit with additional ezDNase treatment according to the manufacturer's instructions (Thermo Fisher Scientific). qRT–PCR was performed using 0.5 μl of cDNA, predesigned TaqMan probes, and TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific), assayed on a LightCycler 480 in 384 well format (Roche). Assays were run in parallel with the loading controls Hprt and Ubc, previously validated to remain constant in the tested assay conditions. Subsequent experiments using ferret epithelial cells were performed using the same methodology.

Human lung tissues. Human samples were obtained under a protocol approved by the Partners Human Research Committee (IRB #2012P001079) and by Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (IRB #1603505962A005).

Single-molecule fluorescent in situ hybridization (smFISH). Segments of human bronchus were flash frozen by immersion in liquid nitrogen and embedded in μOCT and 4-μM sections were collected. RNAScope Multiplex Fluorescent Kit (Advanced Cell Diagnostics) was used per manufacturer's recommendations, and confocal imaging was carried out as described above.

Transwell cultures. Cells were cultured and expanded in complete SAGM (small airway epithelial cell growth medium; Lonza, CC-3118) containing TGF-β/BMP4/WNT antagonist cocktails and 5 μM ROCK inhibitor Y-27632 (Selleckbio, S1049). To initiate air–liquid interface (ALI) cultures, airway basal stem cells were dissociated from mouse tracheas and seeded onto transwell membranes. After reaching confluence, medium was removed from the upper chamber. Mucocilliary differentiation was performed with PneumaCult-ALI Medium (StemCell, 05001). Differentiation of airway basal stem cells on an ALI was followed by directly visualizing beating cilia in real time after 10–14 days.

Once air–liquid cultures were fully differentiated, as indicated by beating cilia, treatment cultures were supplemented with 25 ng/ml of recombinant murine IL-13 (Peprotech-stock diluted in water and used fresh) diluted in PneumaCult-ALI Medium, whereas control cultures received an equal volume of water for 72 h. After treatment, whole ALI wells were fixed in 4% PFA, immunostained in whole mount using the same buffers and imaged with a confocal microscope as described above.

Airway surface physiologic parameters. Epithelia derived from *Foxi1*(KO) mice (wild type, heterozygous knockout and homozygous knockout genotypes)

were grown as ALI cultures in transwells as described above and μOCT, particle-tracking microrheology, airway surface pH measurements, and equivalent current (I_{eq}) assays were used to characterize their physiological parameters as described below.

Micro-optical coherence tomography. μOCT was performed as previously described[34,37,48]. In brief, airway surface liquid (ASL) depth and ciliary beat frequency (CBF) were directly assessed via cross-sectional images of the airway epithelium with high resolution (~1 μm) and high acquisition speed (20,480 Hz A-line rate resulting in 40 frames per s at 512 A-line per frame). Quantitative analysis of images was performed in ImageJ[49]. To establish CBF, custom code in Matlab (Mathworks) was used to quantify Fourier analysis of the reflectance of beating cilia. ASL depth was characterized directly by geometric measurement of the respective layers.

Particle-tracking microrheology. Mucus viscosity was measured following the described method[50].

Airway surface pH. A small probe was used to measure airway surface pH as described[35].

Equivalent current assay. Equivalent current assay on mouse ALI was carried out as described[51] with these changes: benzamil was used at 20 μM and CFTR activation was done only with 10 μM forskolin.

Transcriptional activation of *Foxi1* in ferret basal cell cultures. For lentivirus production and transduction, HEK 293T cells were cultured in 10% FBS, 1% penicillin/streptomycin in DMEM. Cells were seeded at ~30% confluency, and were transfected the next day at ~90% confluency. For each flask, 22 μg of plasmid containing the vector pLent-dCas9-VP64 blasticidin or pLent-MS2-p65-HSF1 hygromycin, 16 μg of psPAX2, and 7 μg pMD2 (VSV-G) were transfected using calcium phosphate buffer[52]. The day after transfection, culture medium was removed and replaced with 2% FBS-DMEM and incubated for 24 h. Lentivirus supernatant was collected 48 h after transfection and centrifuged at 5000 r.p.m. for 5 min. Lentivirus was filtered with a 0.45-μm PVDF filter, concentrated by Lenti X concentrator (Takara), aliquoted and stored at −80 °C. Ferret basal cells were cultured in Pneumacult-Ex with medium supplemented with Pneumacult-Ex and supplemented with hydrocortisone and 1% penicillin/streptomycin and passaged at a 1:5 ratio. Cells were incubated with lentivirus for 24 h in growth medium. At 72 h, selection was initiated (10 μg/ml blasticidin, 50 μg/ml hygromycin). Selection was performed for 14 days for Hygromycin and Blasticidin with media changes every 24 h.

To generate small guide (sg)RNA for transcriptional activation of Foxi1 in ferret cells, gBlocks were synthesized from IDT and included all components necessary for sgRNA production, namely: T7 promoter, *Foxi1* target-specific sequence, guide RNA scaffold, MS2 binding loop and termination signal. gBlocks were PCR amplified and gel purified. PCR products were used as the template for in vitro transcription using MEGAshortscript T7 kit (Ambion). All sgRNAs were purified using MegaClear Kit (Ambion) and eluted in RNase-free water.

Foxi1 sgRNA was reverse transfected using Lipofectamine RNAiMAX Transfection Reagent (Life Science) into ferret basal cells that stably express dCas9-VP64 fusion protein and MS2-p65-HSF1 fusion protein. For the 0.33-cm^2 ALI inserts, (1 μg) sgRNA and Lipofectamine RNAiMAX were diluted in 50 μl of Opti-MEM. The solution was gently mixed, dispensed into insert and incubated for 20–30 min at room temperature. Next, 300,000 cells were suspended in 150 μl Pneumacult-Ex plus medium and incubated for 24 h at 37 °C in a 5% CO_2 incubator.

Short-circuit current measurements of CFTR-mediated chloride transport in ferret. Polarized ferret basal cells with activated *Foxi1* expression as well as matched mock transfection controls (without DNA) were grown in ALI, and after three weeks short-circuit current (I_{sc}) measurements were performed as previously described[53]. The basolateral chamber was filled with high-chloride HEPES-buffered Ringer's solution (135 mM NaCl, 1.2 mM $CaCl_2$, 1.2 mM $MgCl_2$, 2.4 mM KH_2PO_4, 0.2 mM K_2HPO_4, 5 mM HEPES, pH 7.4). The apical chamber received a low-chloride HEPES-buffered Ringer's solution containing a 135-mM sodium gluconate substitution for NaCl. I_{sc} was recorded using Acquire & Analyze software (Physiologic Instruments) after clamping the transepithelial voltage to zero. The following antagonists and agonists were sequentially added into the apical chamber: amiloride (100 μM) to block ENaC channels, apical DIDS (100 μM) to block calcium-activated chloride channels, forskolin (100 μM) and IBMX (100 μM) to activate CFTR, and GlyH 101(100 μM) to block CFTR.

Pre-processing of 3′ droplet-based scRNA-seq data. Demultiplexing, alignment to the mm10 transcriptome and UMI-collapsing were performed using the Cellranger toolkit (version 1.0.1, 10X Genomics). For each cell, we quantified the number of genes for which at least one read was mapped, and then excluded all cells with fewer than 1,000 detected genes. Expression values $E_{i,j}$ for gene i in cell j were calculated by dividing UMI count values for gene i by the sum of the UMI counts in cell j, to normalize for differences in coverage, and then multiplying by 10,000 to create TPM-like values, and finally calculating $\log_2$(TPM+1) values.

Selection of variable genes was performed by fitting a generalized linear model to the relationship between the squared coefficient of variation (CV) and the mean expression level in log–log space, and selecting genes that significantly deviated ($P < 0.05$) from the fitted curve, as previously described[54].

Both prior knowledge and our data show that different cell types have markedly differing abundances in the trachea. For example, 3,845 of the 7,193 cells (53.5%) in the droplet-based dataset were eventually identified as basal cells, whereas only 26 were ionocytes (0.42%). This makes conventional batch correction difficult, as, because of random sampling effects, some batches may have very few (or even zero) of the rarest cells (Extended Data Fig. 1b). To avoid this problem and simultaneously identify maximally discriminative genes, we performed an initial round of clustering on the set of variable genes described above, and identified a set of 1,380 cell-type-specific genes (FDR <0.01), with a minimum $\log_2$ fold-change of 0.25. In addition, we performed batch correction within each identified cluster, which contained only transcriptionally similar cells, ameliorating problems with differences in abundance. Batch correction was performed (only on these 1,380 genes) using ComBat[55] as implemented in the R package sva[56] using the default parametric adjustment mode. The output was a corrected expression matrix, which was used as input to further analysis.

Pre-processing of plate-based scRNA-seq data. BAM files were converted to merged, de-multiplexed FASTQ files using the Illumina Bcl2Fastq software package v.2.17.1.14. Paired-end reads were mapped to the UCSC mm10 mouse transcriptome using Bowtie[57] with parameters '-q–phred33-quals -n 1 -e 99999999 -l 25 -I 1 -X 2000 -a -m 15 -S -p 6', which allows alignment of sequences with one mismatch. Expression levels of genes were quantified as transcript-per-million (TPM) values by RSEM[58] v.1.2.3 in paired-end mode. For each cell, we determined the number of genes for which at least one read was mapped, and then excluded all cells with fewer than 2,000 detected genes. We then identified highly variable genes as described above.

Dimensionality reduction by PCA and *t*-SNE. We restricted the expression matrix to the subsets of variable genes and high-quality cells noted above, and values were centred and scaled before input to PCA, which was implemented using the R function 'prcomp' from the 'stats' package for the plate-based dataset. For the droplet-based dataset, we used a randomized approximation to PCA, implemented using the 'rpca' function from the 'rsvd' R package, with the parameter k set to 100. This low-rank approximation is several orders of magnitude faster to compute for very wide matrices. After PCA, significant principal components were identified using a permutation test as previously described[59], implemented using the 'permutationPA' function from the 'jackstraw' R package. Because of the presence of extremely rare cells in the droplet-based dataset (as described above), we used scores from 10 significant principal components using scaled data, and 7 significant principal components using unscaled data. Only scores from these significant PCs were used as the input to further analysis.

For visualization purposes only (and not for clustering), dimensionality was further reduced using the Barnes–Hut approximate version of *t*-SNE[60,61]. This was implemented using the 'Rtsne' function from the 'Rtsne' R package using 20,000 iterations and a perplexity setting of 10 and 75 for plate- and droplet-based, respectively. Scores from the first n principal components were used as the input to *t*-SNE, in which n was 11 and 12 for plate- and droplet-based data, respectively, determined using the permutation test described above.

Excluding immune, mesenchymal cells and suspected doublets. Although cells were sorted using EpCAM before scRNA-seq, 1,873 contaminating cells were observed in the initial droplet dataset, and were comprised of: 91 endothelial cells expressing *Egfl7*, *Sh3gl3* and *Esam*; 229 macrophages expressing MHCII (*H2-Ab1*, *H2-Aa*, *Cd74*), *C1qa*, and *Cd68*; and 1,553 fibroblasts expressing high levels of collagens (*Col1a1*, *Col1a2* and *Col3a1*). Each of these cell populations was identified by an initial round of unsupervised clustering (density-based clustering of the *t*-SNE map using 'dbscan'[75] from the R package 'fpc') as they formed extremely distinct clusters, and then removed. In the case of the pulse-seq dataset, the initial clustering step removed a total of 532 dendritic cells identified by high expression of *Ptprc* and *Cd83*. In addition, 20 other cells were outliers in terms of library complexity, which could possibly correspond to more than one individual cell per sequencing library, or 'doublets'. As a conservative precaution, we removed these 20 possible doublet cells with over 3,700 genes detected per cell.

***k* nearest neighbour-graph based clustering.** To cluster single cells by their expression profiles, we used unsupervised clustering, based on the Infomap community-detection algorithm[62], following approaches recently described for single-cell CyTOF data[63] and scRNA-seq[64]. We constructed a k nearest-neighbour graph using, for each pair of cells, the Euclidean distance between the scores of significant principal components as the metric.

The number k of nearest neighbours was chosen in a manner roughly consistent with the size of the dataset, and set to 25 and 150 for plate- and droplet-based data, respectively. For sub-clustering of rare cell subsets, we used $k = 100$, 50, 50 and 20 for tuft cells, neuroendocrine cells, ionocytes and goblet cells, respectively.

The k nearest neighbour graph was computed using the function 'nng' from the R package 'cccd' and was then used as the input to Infomap[62], implemented using the 'infomap.community' function from the 'igraph' R package.

Detected clusters were mapped to cell types using known markers for tracheal epithelial subsets. In particular, because of the large proportion of basal and club cells, multiple clusters expressed high levels of markers for these two types. Accordingly, we merged nine clusters expressing the basal gene score above a median $\log_2(\text{TPM}+1) > 0$, and seven clusters expressing the club gene score above median $\log_2(\text{TPM}+1) > 1$. Calculation of a ciliated cell gene score showed only a single cluster with non-zero median expression, so no further merging was performed. This resulted in seven clusters, each corresponding 1-to-1 with a known airway epithelial cell type, with the exception of the ionocyte cluster, which we show represents a novel subset.

Rare cells (tuft, neuroendocrine, ionocyte and goblet) were sub-clustered to examine possible heterogeneity of mature types (Fig. 4, Extended Data Fig. 8). In each case, cells annotated as each type from the initial 3′ droplet-based dataset (Fig. 1b, Extended Data Fig. 1d) were combined with the corresponding cells from the pulse-seq dataset (Fig. 3b, Extended Data Fig. 6a) before sub-clustering. In the case of goblet cells, sub-clustering the combined 468 goblet cells ($k = 20$, above) partitioned the data into 7 groups, two of which expressed the novel goblet cell marker *Gp2* (Fig. 1d) at high levels (median $\log_2(\text{TPM}+1) > 1$). These two groups were annotated as mature goblet-1 and goblet-2 cells (Extended Data Fig. 8f–j), while the five groups were merged and annotated as immature goblet cells. No cluster merging was performed for sub-clustering of tuft, neuroendocrine or ionocytes.

Differential expression and cell-type signatures. To identify maximally specific genes for cell-types, we performed differential expression tests between each pair of clusters for all possible pairwise comparisons (larger clusters—basal, club, ciliated cells—were down-sampled to 1,000 cells). Then, for a given cluster, putative signature genes were filtered using the maximum FDR Q value and ranked by the minimum $\log_2$ fold-change (across the comparisons). This is a stringent criterion because the minimum fold-change and maximum Q value represent the weakest effect size across all pairwise comparisons. Cell-type signature genes for the initial droplet based scRNA-seq data (Fig. 1c, Supplementary Tables 1) were obtained using a maximum FDR of 0.05 and a minimum $\log_2$ fold-change of 0.5.

Where fewer cells were available, as is the case for full-length plate-based scRNA-seq data (Extended Data Fig. 3b, Supplementary Table 2) or for subtypes within cell-types (Fig. 3c, Extended Data Fig. 8c), a combined P value across the pairwise tests for enrichment was computed using Fisher's method (a more lenient criterion) and a maximum FDR Q value of 0.001 was used, along with a cut-off of minimum $\log_2$ fold-change of 0.1 for tuft and goblet cell subsets (Fig. 3c, Extended Data Fig. 8c, Supplementary Table 8). Marker genes were ranked by minimum $\log_2$ fold-change. Differential expression tests were carried using a two part 'hurdle' model to control for both technical quality and mouse-to-mouse variation. This was implemented using the R package MAST[76], and P values for differential expression were computed using the likelihood-ratio test. Multiple hypothesis testing correction was performed by controlling the false discovery rate[77] using the R function 'p.adjust'.

Scoring cells using signature gene sets. To obtain a score for a specific set of n genes in a given cell, a 'background' gene set was defined to control for differences in sequencing coverage and library complexity. The background gene set was selected for similarity to the genes of interest in terms of expression level. Specifically, the $10n$ nearest neighbours in the 2D space defined by mean expression and detection frequency across all cells were selected. The signature score for that cell was then defined as the mean expression of the n signature genes in that cell, minus the mean expression of the $10n$ background genes in that cell.

Assigning cell-type specific transcription factors, G-protein-coupled receptors and genes associated with asthma. A list of all genes annotated as transcription factors in mice was obtained from AnimalTFDB[67], downloaded from http://www.bioguo.org/AnimalTFDB/BrowseAllTF.php?spe=Mus_musculus. The set of G-protein-coupled receptors (GPCRs) was obtained from the UniProt database, downloaded from http://www.uniprot.org/uniprot/?query=family%3A%22g+protein+coupled+receptor%22+AND+organism%3A%22Mouse+%5B10090%5D%22+AND+reviewed%3Ayes&sort=score. To map from human to mouse gene names, human and mouse orthologues were downloaded from Ensembl latest release 86 at http://www.ensembl.org/biomart/martview, and human and mouse gene synonyms from NCBI (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO/Mammalia/).

Cell-type enriched transcription factors and GPCRs were then identified by intersecting the list of genes enriched in to each cell type with the lists of transcription factors and GPCRs defined above. Cell-type enriched transcription factors (Fig. 1e) and GPCRs (Extended Data Fig. 8a) were defined using the 3′ droplet-based and full-length plate-based datasets, respectively, as those with a minimum $\log_2$ fold-change of 0.1 and a maximum FDR of 0.001, retaining a maximum of 10 genes per cell type in Fig. 1e; complete lists are provided in Supplementary Table 4.

Gene set or pathway enrichment analysis. Gene ontology (GO) analysis of enriched pathways in Krt13$^+$ hillocks (Extended Data Fig. 3d) was performed using the 'goseq' R package[68], using significantly differentially expressed genes (FDR < 0.05) as target genes, and all genes expressed with $\log_2(\text{TPM}+1) > 3$ in at least 10 cells as background. For pathway and gene sets, we used a version of MSigDB[69] with mouse orthologues, downloaded from http://bioinf.wehi.edu.au/software/MSigDB/. Association of principal components with cell-types (Extended Data Fig. 7a, b) was computed using the gene-set enrichment analysis (GSEA) algorithm implemented using the 'fgsea' package in R[65]. Genes that are involved in leukotriene biosynthesis and taste transduction pathways (Fig. 4c) were identified using KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO pathways. Specifically, genes in KEGG pathway 00590 (arachidonic acid metabolism) or GO terms 0019370 (leukotriene biosynthetic process) or 0061737 (leukotriene signalling pathway) were annotated as leukotriene-synthesis-associated, whereas genes in KEGG pathway 04742 (taste transduction) were annotated as taste-transduction-associated. To identify statistical enrichment of these taste and leukotriene pathways in tuft-1 and tuft-2 subtypes, respectively, the hypergeometric probability of the overlap between the marker genes for each subset (Supplementary Table 8) and the gene sets was directly calculated using the R function 'fisher.test'.

Statistical analysis of proximodistal mucous metaplasia. For the analysis in Fig. 2h, i, the extent of goblet cell hyperplasia was assessed using counts of Muc5ac$^+$ goblet cells, normalized to counts of EGFP$^+$ ciliated cells. To quantify differences in the count values between the samples in different conditions ($n = 6$, *Foxj1*-EGFP mice), we fit a negative binomial regression using the 'glm.nb' function from the 'MASS' package in R. Pairwise comparisons between means for each condition were computed using post hoc tests and P values were adjusted for multiple comparisons using Tukey's HSD, implemented using the function 'pairs' from the 'emeans' package in R.

Lineage inference using diffusion maps. We restricted our analysis to the 6,848 cells in basal, club or ciliated cell clusters (95.2% of the 7,193 cells in the initial droplet dataset), because it was unlikely that rare cells (for example, neuroendocrine, tuft, goblet and ionocyte cells) in transitional states would be sufficiently densely sampled. Next, we selected highly variable genes among these three cell subsets as described above, and performed dimensionality reduction using the diffusion map approach[66]. In brief, a cell–cell transition matrix was computed using the Gaussian kernel in which the kernel width was adjusted to the local neighbourhood of each cell, following the previously described approach[70]. This matrix was converted to a Markovian matrix after normalization. The right eigenvectors $v_i (i = 0,1,2,3,\ldots)$ of this matrix were computed and sorted in the order of decreasing eigenvalues $\lambda_i (i = 0,1,2,3,\ldots)$, after excluding the top eigenvector v_0, corresponding to $\lambda_0 = 1$ (which reflects the normalization constraint of the Markovian matrix). The remaining eigenvectors $v_i (i = 1,2,\ldots)$ define the diffusion map embedding and are referred to as diffusion components ($DC_k (k = 1,2,\ldots)$). We noticed a spectral gap between λ_3 and λ_4, and hence retained DC_1 – DC_3 for further analysis.

To extract the edges of this manifold, along which cells transition between states (Fig. 2a), we fit a convex hull using the 'convhulln' from the 'geometry' R package. To identify edge-associated cells, any cell within $d < 0.1$ of an edge of the convex hull (in which d is the Euclidean distance in diffusion space) is assigned to that edge.

To identify cells associated with the *Krt4*$^+$/*Krt13*$^+$ population, we used unsupervised partitioning around medoids (PAM) clustering of the cells in diffusion space with the parameter $k = 4$. Edge-association of genes (or transcription factors, Supplementary Table 7) was computed as the autocorrelation (lag = 25), implemented using the 'acf' function from the 'stats' R package. Empirical P values for each edge-associated gene were assessed using a permutation test (1,000 bootstrap iterations), using the autocorrelation value as the test statistic.

Genes were placed in pseudotemporal order by splitting the interval into 30 bins from 'early' to 'late', and assigning each gene the bin with the highest mean expression. These data were smoothed using loess regression and then visualized as heat maps (Extended Data Fig. 5).

Pulse-seq data analysis. For the much larger pulse-seq dataset (66,265 cells), we used a very similar, but more scalable, analysis pipeline, with the following modifications. Alignment and UMI collapsing was performing using the Cellranger toolkit (version 1.3.1, 10X Genomics). $\text{Log}_2(\text{TPM}+1)$ expression values were computed using Rcpp-based function in the R package 'Seurat' (v2.2). We also used an improved method of identifying variable genes. Rather than fitting the mean–CV^2 relationship, a logistic regression was fit to the cellular detection fraction (often referred to as α), using the total number of UMIs per cell as a predictor. Outliers from this curve are genes that are expressed in a lower fraction of cells than would be expected given the total number of UMIs mapping to that gene, that is, cell-type or state-specific genes. We used a threshold of deviance < -0.25, producing a set of 708 variable genes. We restricted the expression matrix

to this subset of variable genes and values were centred and scaled—while 'regressing out'[71] technical factors (number of genes detected per cell, number of UMIs detected per cell and cell-cycle score) using the 'ScaleData' function before input to PCA, implemented using 'RunPCA' in Seurat. After PCA, significant principal components were identified using the knee in the scree plot, which identified 10 significant principal components. Only scores from these significant PCs were used as the input to nearest-neighbour based clustering and *t*-SNE, implemented using the 'FindClusters' (resolution parameter $r=1$) and 'RunTSNE' (perplexity $P=25$) methods, respectively, from the 'Seurat' package.

Once again, owing to their abundance, the populous basal, club and ciliated cells were spread across several clusters, which were merged using the strategy described above: 19 clusters expressing the basal score above mean $\log_2(\text{TPM}+1) > 0$, 12 expressing the club score above mean $\log_2(\text{TPM}+1) > -0.1$, and 2 clusters expressing the ciliated signature above were merged to construct the basal, club and ciliated subsets, respectively. Goblet cells were not immediately associated with a specific cluster, however, cluster 13 (one of those merged into the club cluster) expressed significantly elevated levels of goblet markers *Tff2* and *Gp2* ($P < 10^{-10}$, LRT). Sub-clustering this population (resolution parameter $r=1$) revealed 6 clusters, of which two expressed the goblet score constructed using the top 25 goblet cell marker genes (Supplementary Table 1) above mean $\log_2(\text{TPM}+1) > 1$, which were merged and annotated as goblet cells. To identify the *Krt4*$^+$/*Krt13*$^+$ hillock-associated club cells, the remaining 17,700 club cells were re-clustered (resolution parameter $r=0.2$) into 5 clusters, of which one expressed much higher levels ($P < 10^{-10}$ in all cases) of *Krt4*, *Krt13* and a hillock score constructed using the top 25 hillock marker genes (Supplementary Table 6), this cluster was annotated as 'hillock-associated club cells'.

Estimating lineage-labelled fraction for pulse-seq and conventional lineage tracing. For any given sample (here, mouse) the certainty in the estimate of the proportion of labelled cells increases with the number of cells obtained; the more cells, the higher the precision of the estimate. Estimating the overall fraction of labelled cells (from conventional lineage tracing, Fig. 3f, Extended Data Figs. 4, 6; or pulse-seq lineage tracing, Fig. 3, Extended Data Fig. 6) on the basis of the individual estimates from each mouse is analogous to performing a meta-analysis of several studies, each of which measures a proportion of the population; studies with greater power (higher n) carry more information, and should influence the overall estimate more, whereas low-n studies provide less information and should not have as much influence. Generalized linear mixed models provide a framework to obtain an overall estimate in this manner[72]. Accordingly, we implemented a fixed effects logistic regression model to compute the overall estimate and 95% confidence interval using the function 'metaprop' from the R package 'meta'[73].

Testing for difference in labelled fraction for pulse-seq and conventional lineage tracing. To assess the significance of changes in the labelled fraction of cells in different conditions, we used a negative binomial regression model of the counts of cells at each time-point, controlling for variability amongst biological (mouse) replicates. For each cell type, we model the number of lineage-labelled cells detected in each analysed mouse as a random count variable using a negative binomial distribution. The frequency of detection is modelled by using the natural log of the total number of cells of that type profiled in a given mouse as an offset. The time point of each mouse (0, 30 or 60 days post tamoxifen) is provided as a covariate. The negative binomial model was fit using the R command 'glm.nb' from the 'MASS' package. The *P* value for the significance of the change in labelled fraction size between time-points was assessed using a likelihood-ratio test, computing using the R function 'anova'.

Estimating turnover rate using quantile regression. Given the relatively few samples ($n=9$ mice) with which to model the rate of new lineage-labelled cells, we used the more robust quantile regression[74], which models the conditional median (rather than the conditional mean, as captured by least-squares linear regression, which can be sensitive to outliers). The fraction of labelled cells in each mouse was modelled as a function of days post tamoxifen (Extended Data Fig. 6b) using the function 'rq' from the R package 'quantReg'. Significance of association between increasing labelled fraction and time were computing using Wald tests implemented with the 'summary.rq' function, while tests comparing the slopes of fits were conducted using 'anova.rq'.

Statistics. Blinding was used for data analysis including the genotype of mouse samples for qRT-PCR expression studies, electrophysiology studies, and characterization of physiologic parameters at the epithelial surface (pH, ASL, mucus, CBF, viscosity).

Statistical hypothesis testing. With the exception of the LRT, which is one-tailed, all tests used were two-tailed, and exact *P* values are reported, except where they are below the threshold of numerical precision (2.22×10^{-16}).

Statistical analysis of qRT–PCR data. $\Delta\Delta C_T$ values were generated by normalization to the average of loading controls *Hprt* and *Ubc*, followed by comparison to wild-type samples. Statistical analysis was performed at the ΔC_T stage. For single comparisons, all datasets passed the Shapiro–Wilk normality test, which was followed by a post hoc two-tailed *t*-test. For multiple comparisons, all datasets passed the Shapiro–Wilk normality test for equal variance. Data were then tested by two-way ANOVA, with sex as the second level of variance. In a few specific cases, sex trended towards significance, however, not sufficiently to justify separate analysis. Post hoc multiple comparisons to the control group were performed using the Holm–Sidak method. In the single case of Foxi1 KO (Fig. 5e), two heterozygous samples were identified as outliers and removed using a standard implementation of DBscan clustering using the full dataset of all genes assayed using qRT–PCR. These two samples exhibited gene expression closer to full Foxi1 knockouts and were removed from consideration. In all cases, error bars represent the calculated 95% CI.

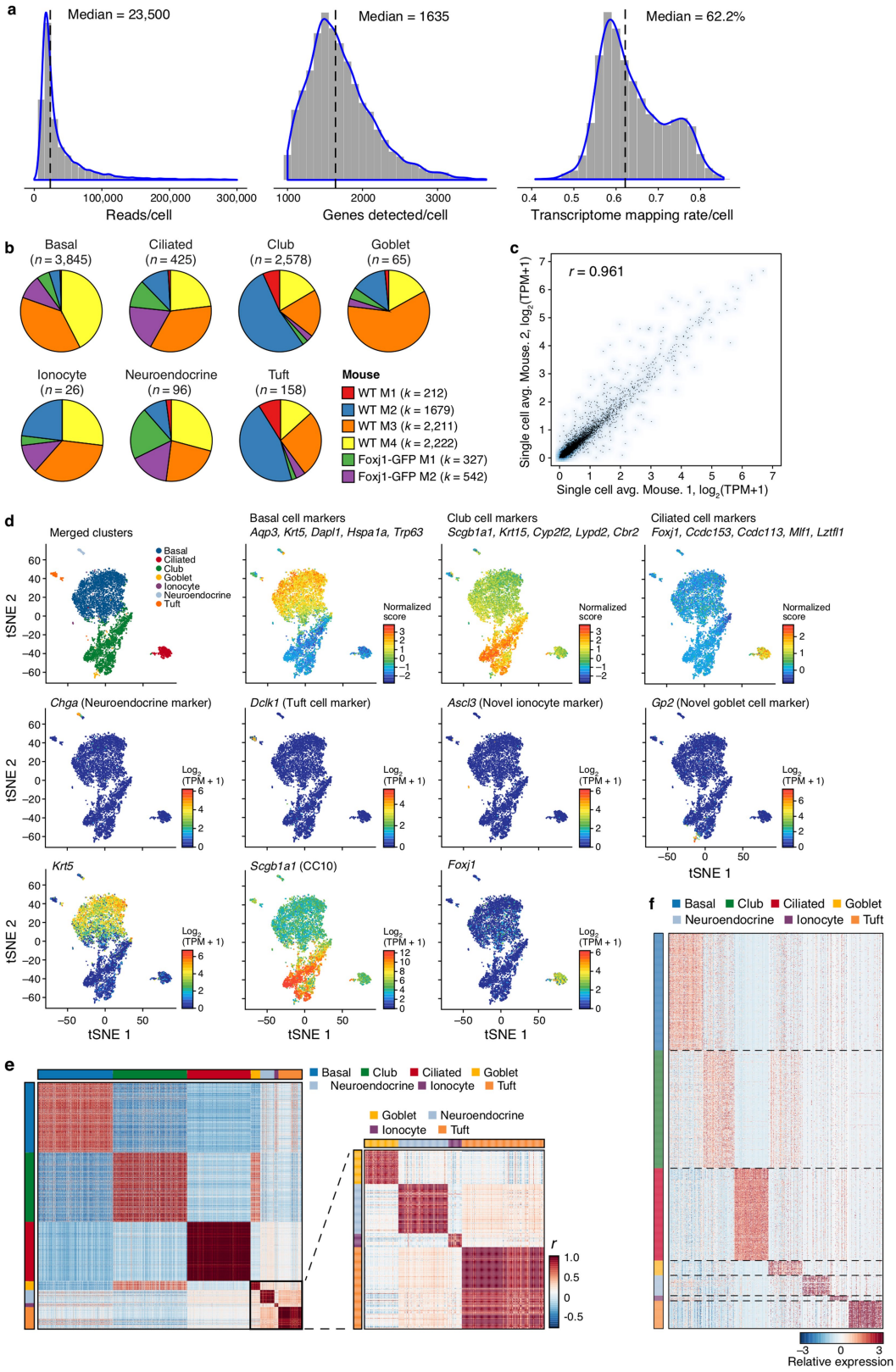
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. R markdown scripts enabling the main steps of the analysis to be performed are available from https://github.com/adamh-broad/single_cell_airway.

Data availability. All data have been deposited in Gene Expression Omnibus under accession code GSE103354 and in the Single Cell Portal (https://portals.broadinstitute.org/single_cell/study/airway-epithelium), and Source Data for Figs. 1–5 is provided with the paper.

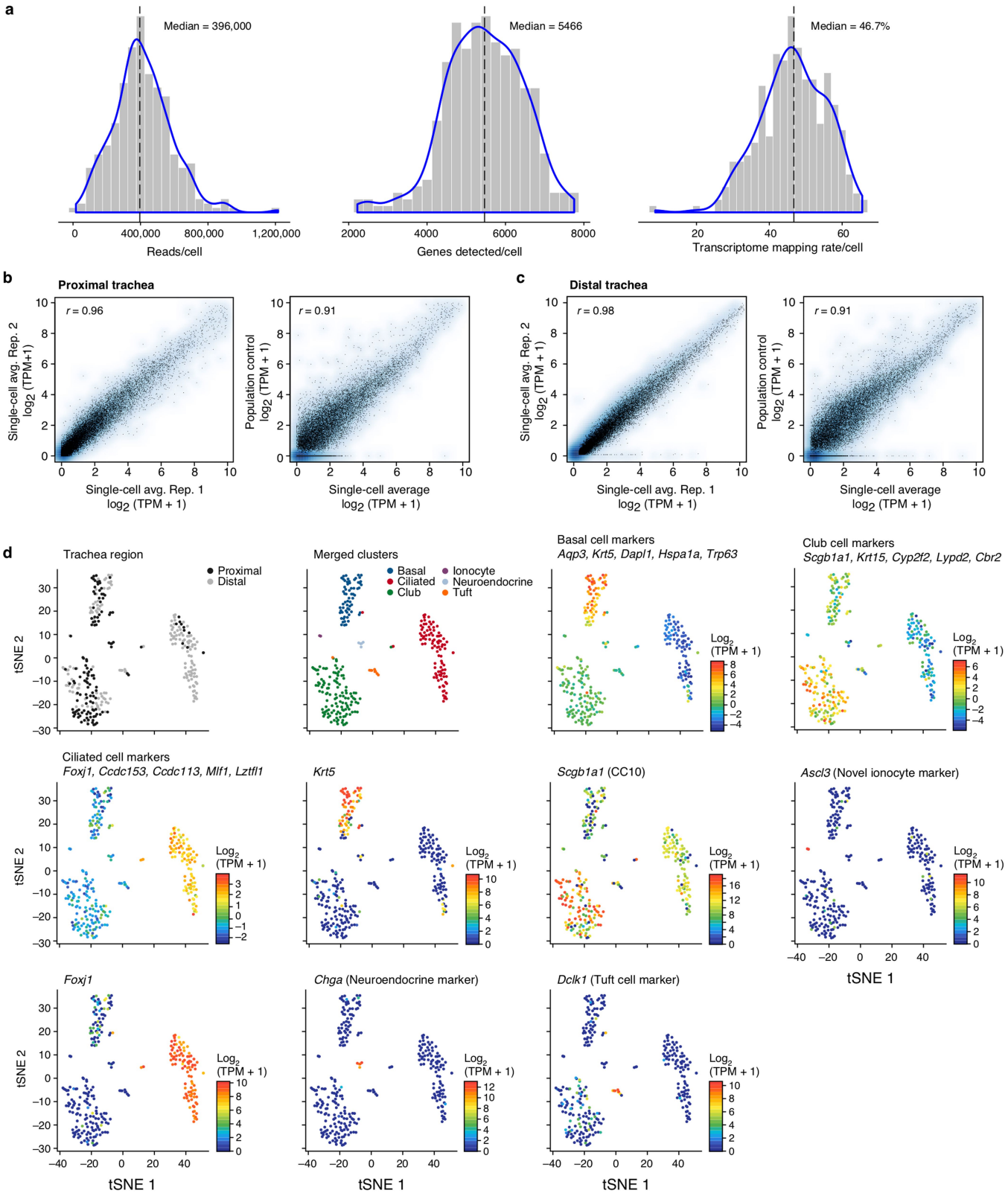
45. Rawlins, E. L. et al. The role of Scgb1a1$^+$ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4**, 525–534 (2009).
46. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis *in vivo*. *Proc. Natl Acad. Sci. USA* **105**, 2415–2420 (2008).
47. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
48. Liu, L. et al. Method for quantitative study of airway functional microanatomy using micro-optical coherence tomography. *PLoS ONE* **8**, e54473 (2013).
49. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
50. Birket, S. E. et al. Combination therapy with cystic fibrosis transmembrane conductance regulator modulators augment the airway functional microanatomy. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **310**, L928–L939 (2016).
51. Mou, H. et al. Dual SMAD signaling inhibition enables long-term expansion of diverse epithelial basal cells. *Cell Stem Cell* **19**, 217–231 (2016).
52. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
53. Yan, Z. et al. Optimization of recombinant adeno-associated virus-mediated expression for large transgenes, using a synthetic promoter and tandem array enhancers. *Hum. Gene Ther.* **26**, 334–346 (2015).
54. Brennecke, P. et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
55. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
56. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
57. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
58. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
59. Buja, A. & Eyuboglu, N. Remarks on parallel analysis. *Multivariate Behav. Res.* **27**, 509–540 (1992).
60. Van Der Maaten, L. Accelerating *t*-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
61. van der Maaten, L. & Hinton, G. Visualizing Data using *t*-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
62. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123 (2008).
63. Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
64. Shekhar, K. et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323 (2016).
65. *fgsea: Fast Gene Set Enrichment Analysis*. (Computer Technologies Laboratory, 2018).
66. Coifman, R. R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: multiscale methods. *Proc. Natl Acad. Sci. USA* **102**, 7432–7437 (2005).
67. Zhang, H.-M. et al. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* **40**, D144–D149 (2012).
68. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
69. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
70. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
71. Buettner, F. et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
72. Gelman, A & Hill, J. *Data Analysis Using Regression and Multilevel/hierarchical Models* (Cambridge Univ. Press, New York, 2007).
73. Schwarzer, G et al. (eds) *Meta-Analysis with R* (Springer, New York, 2015).

74. Koenker, R. & Hallock, K. F. Quantile regression. *J. Econ. Perspect.* **15**, 143–156 (2001).
75. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining* (Simoudis E. et al. eds) 226–231 (AAAI, 1996).
76. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
77. Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. C* **57**, 289–300.

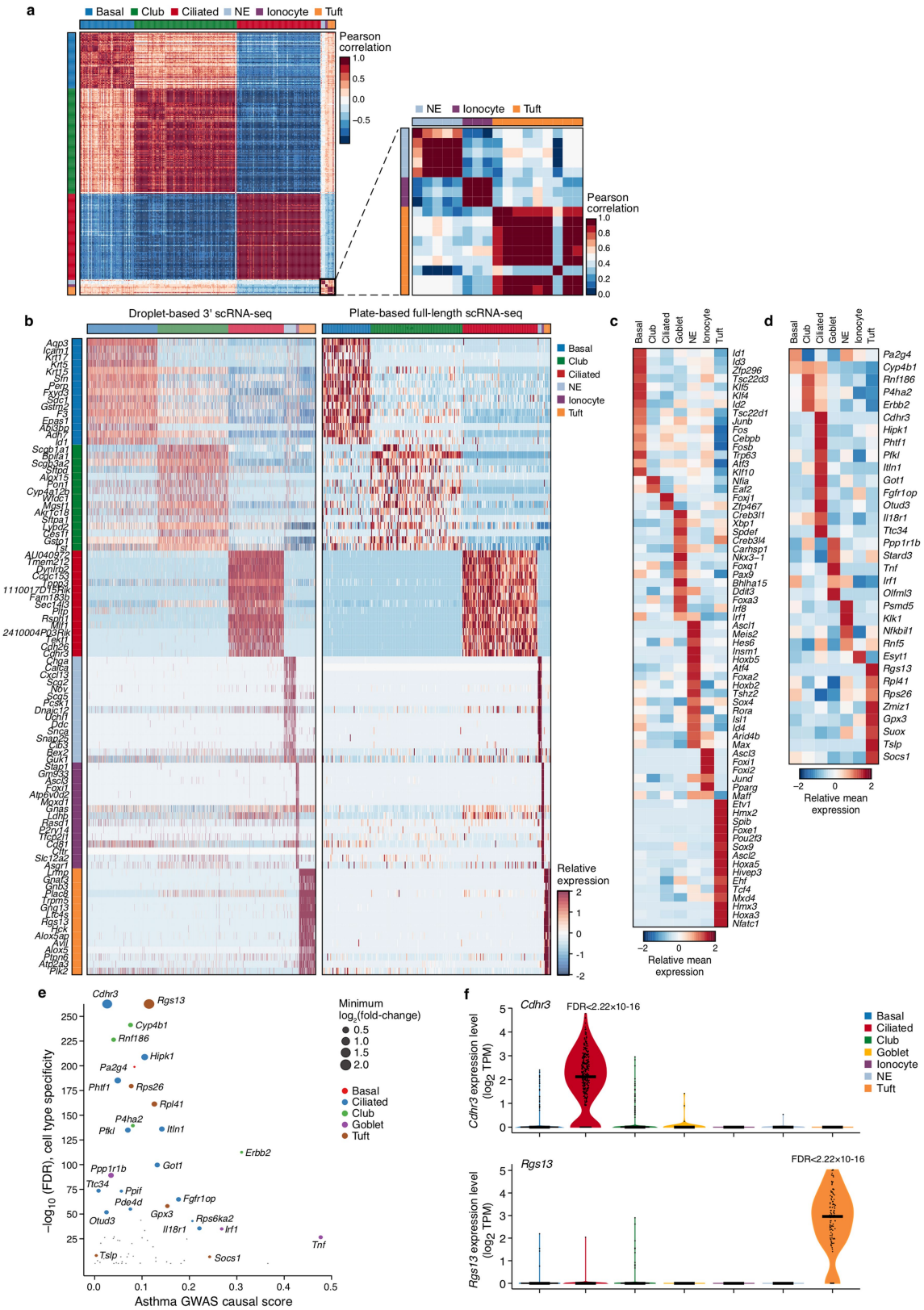


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Identifying tracheal epithelial cell types in 3′ scRNA-seq. a, Quality metrics for the initial droplet-based 3′ scRNA-seq data. Distributions of the number of reads per cell (left), the number of the genes detected with non-zero transcript counts per cell (centre), and the fraction of reads mapping to the mm10 transcriptome per cell (right). Dashed line, median; blue line, kernel density estimate. **b**, Cell type clusters are composed of cells from multiple biological replicates. Fraction of cells in each cluster that originate from a given biological replicate ($n = 6$ mice). Post hoc annotation and number of cells are indicated above each pie chart. All biological replicates contribute to all clusters (except for wild-type mouse 1, which did not contain any of the very rare ionocytes (0.39% of all epithelial cells)), and no significant batch effect was observed. **c**, Reproducibility between biological replicates. Average gene expression values ($\log_2$(TPM+1)) across all cells of two representative 3′ scRNA-seq replicate experiments (r = Pearson correlation coefficient). Blue shading, gene (point) density. **d**, Post hoc cluster interpretation based on the expression of known cell type markers. *t*-SNE of 7,193 scRNA-seq profiles (points), coloured by cluster assignment (top left), by expression ($\log_2$(TPM+1)) of single marker genes, or by mean expression of several marker genes[4] for a particular cell type. **e**, Cell type clusters. Pearson correlation coefficients (r, colour bar) between every pair of 7,193 cells (rows and columns) ordered by cluster assignment. Inset (right), zoom of 288 cells from the rare types. **f**, Gene signatures. Relative expression level (row-wise Z score of $\log_2$(TPM+1) expression values) of cell-type-specific genes (rows) in each epithelial cell (columns). Large clusters (basal, club) are down-sampled to 500 cells.

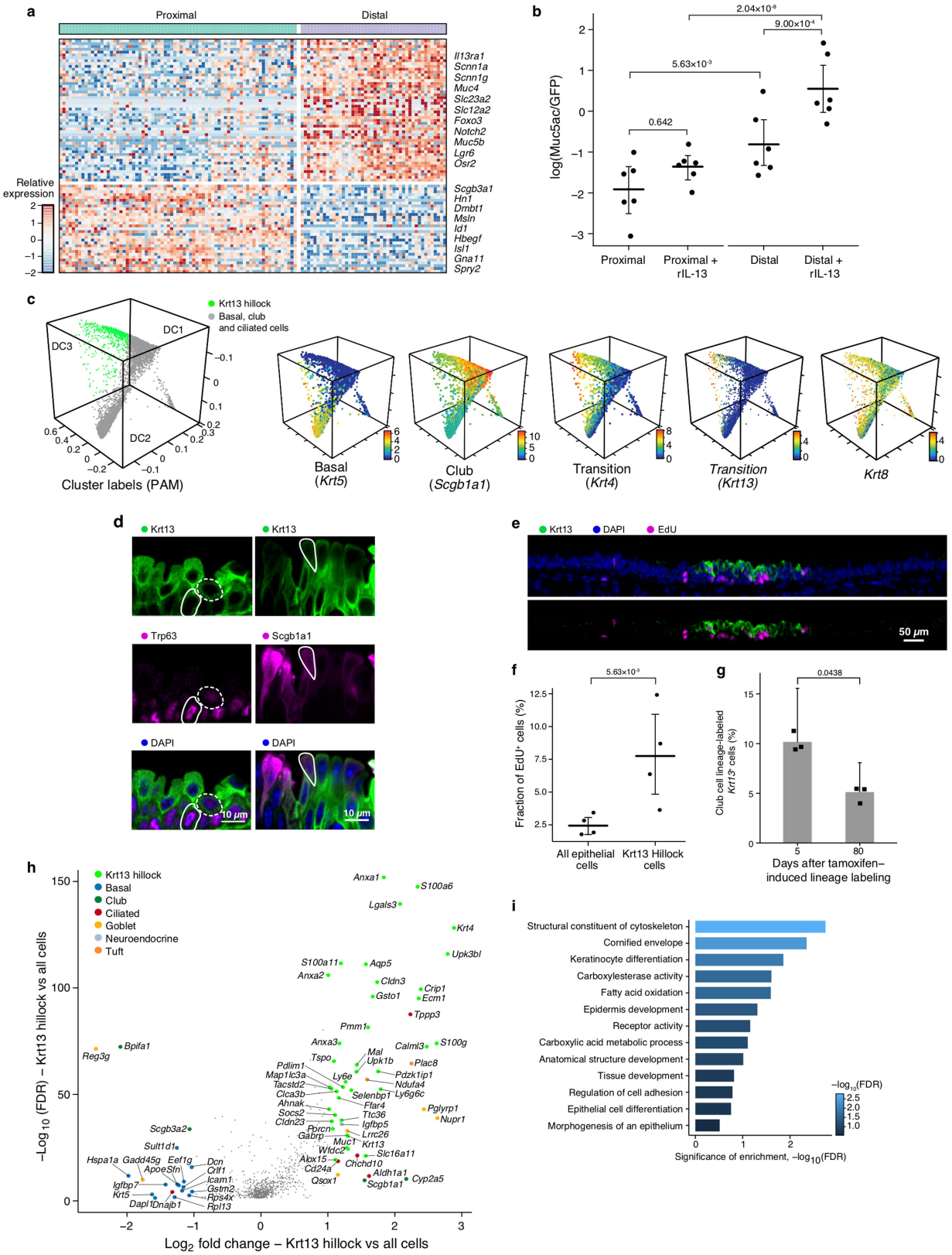


Extended Data Fig. 2 | Identifying tracheal epithelial cell types in full-length scRNA-seq. a, Quality metrics for full-length, plate-based scRNA-seq data. Distributions of the number of reads per cell (left), the number of the genes detected with non-zero transcript counts per cell (centre), and the fraction of reads mapping to the mm10 transcriptome per cell (right). **b**, **c**, High reproducibility between plate-based scRNA-seq data from biological replicates of tracheal epithelial cells. Average expression values ($\log_2$(TPM+1)) in two representative full-length scRNA-seq replicate experiments (left) and in the average of a full-length scRNA-seq dataset (right) and a population control (right) for cells extracted from proximal (**b**) and distal (**c**) mouse trachea. Blue shading: density of genes (points); r = Pearson correlation coefficient. **d**, Post hoc cluster annotation by the expression of known cell-type markers. *t*-SNE of 301 scRNA-seq profiles (points) coloured by region of origin (top left), cluster assignment (top, second from left), or, in the remaining plots, the expression level ($\log_2$(TPM+1)) of single marker genes or the mean expression of several marker genes for a particular cell type. All clusters are populated by cells from both proximal and distal epithelium except rare neuroendocrine cells, which were only detected in proximal experiments (top left).



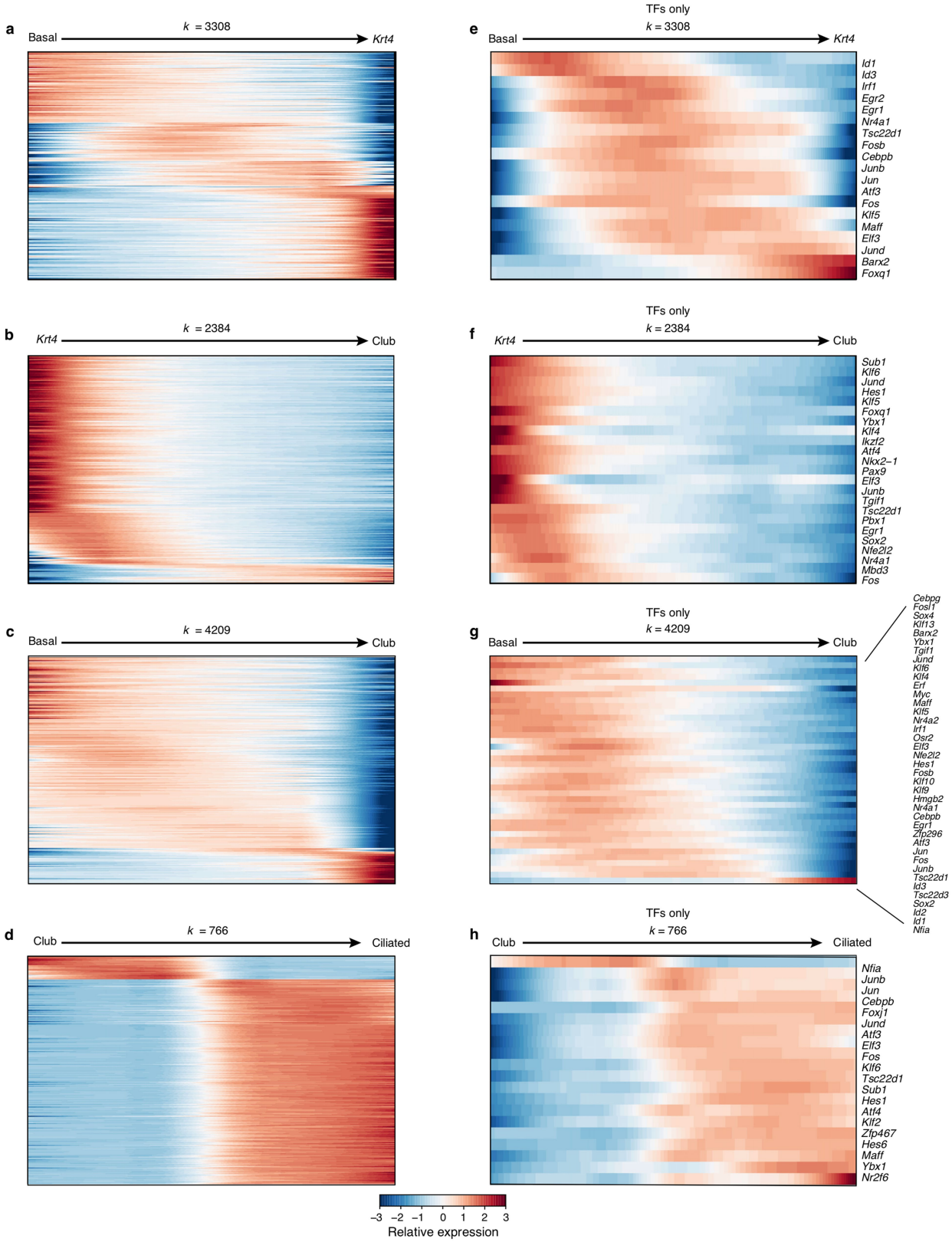
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | High-confidence consensus cell type markers, and cell-type-specific expression of asthma-associated genes. a, Cell type clusters in full-length plate-based scRNA-seq data. Cell–cell Pearson correlation coefficient (*r*) between all 301 cells (individual rows and columns) ordered by cluster assignment (as in Extended Data Fig. 2d). Right, magnified view of 17 cells (black border on left) from the rare types. **b**, High confidence consensus markers. Relative expression level (row-wise *Z* score of mean $\log_2$(TPM+1)) of consensus marker genes (rows, FDR <0.01 in both 3′-droplet and full-length plate-based scRNA-seq datasets; LRT) for each cell type (flanking colour bar) across 7,193 cells in the 3′ droplet data (columns, left) and the 301 cells in the plate-based dataset (columns, right). Top 15 markers shown, complete sets are in Extended Data Fig. 1f, Supplementary Table 3. **c**, Cluster-specific transcription factors in 3′ scRNA-seq data. Mean relative expression (row-wise *Z* score of mean $\log_2$(TPM+1), colour bar) of the top transcription factors (rows) that are enriched (FDR <0.01, LRT, two-sided) in cells (columns) of each cluster. **d**–**f**, Cell-type-specific expression of genes associated with asthma by GWAS. **d**, Relative expression (*Z* score of mean $\log_2$(TPM+1)) of genes that are associated with asthma in GWAS and enriched (FDR <0.01, LRT) for cell-type-specific expression in our 3′ scRNA-seq data. **e**, The significance ($-\log_{10}$(FDR), Fisher's combined *P* value, LRT) and effect size (point size, mean $\log_2$(fold-change)) of cell-type-specific expression and its genetic association strength from GWAS[15] for each gene from **d**. **f**, Distribution of expression levels ($\log_2$(TPM+1)) in the cells in each cluster (*x* axis, colour legend) for two asthma GWAS genes: *Cdhr3* (top; specific to ciliated cells) and *Rgs13* (bottom; specific to tuft cells) FDRs, LRT.

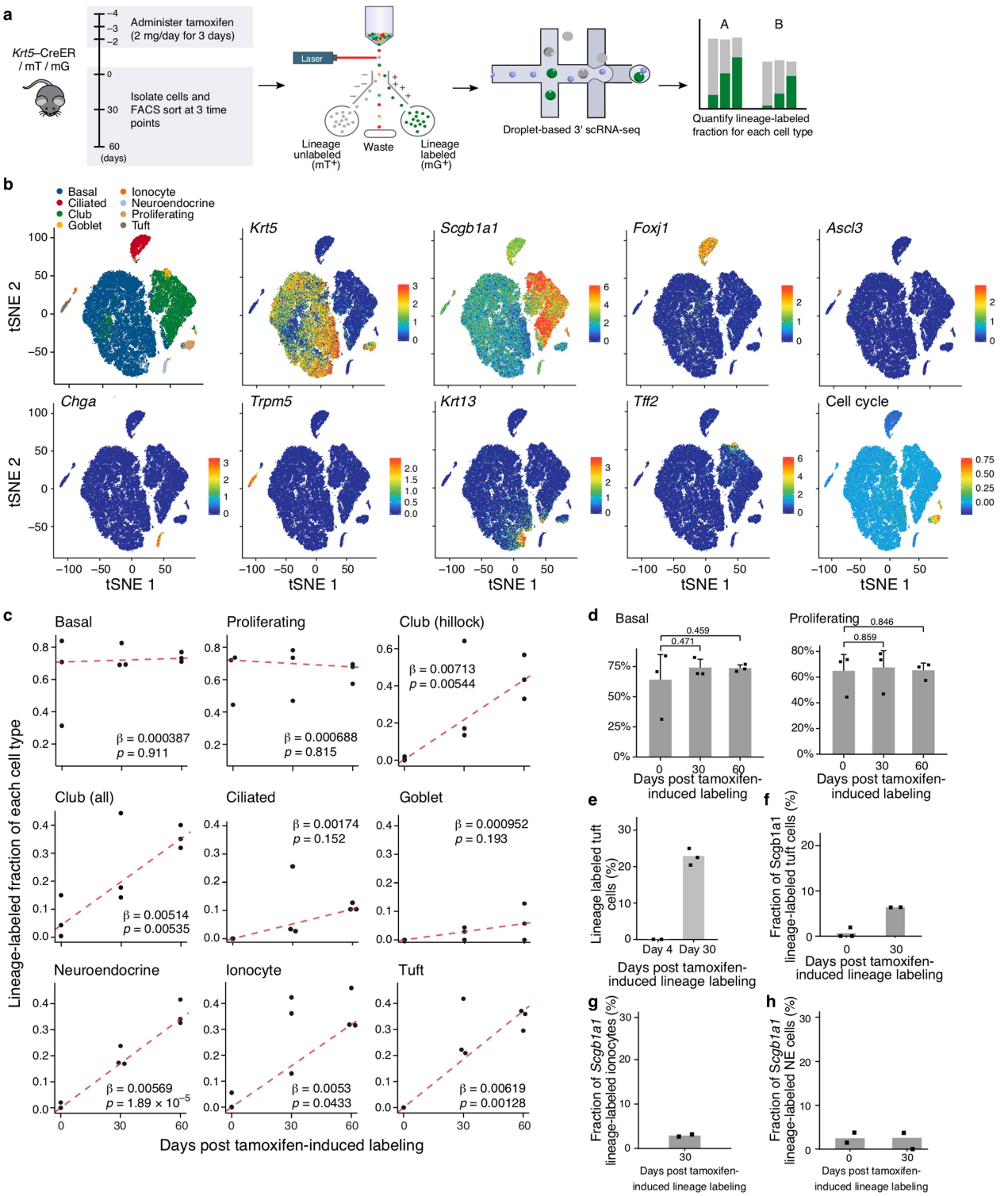


Extended Data Fig. 4 | See next page for caption.

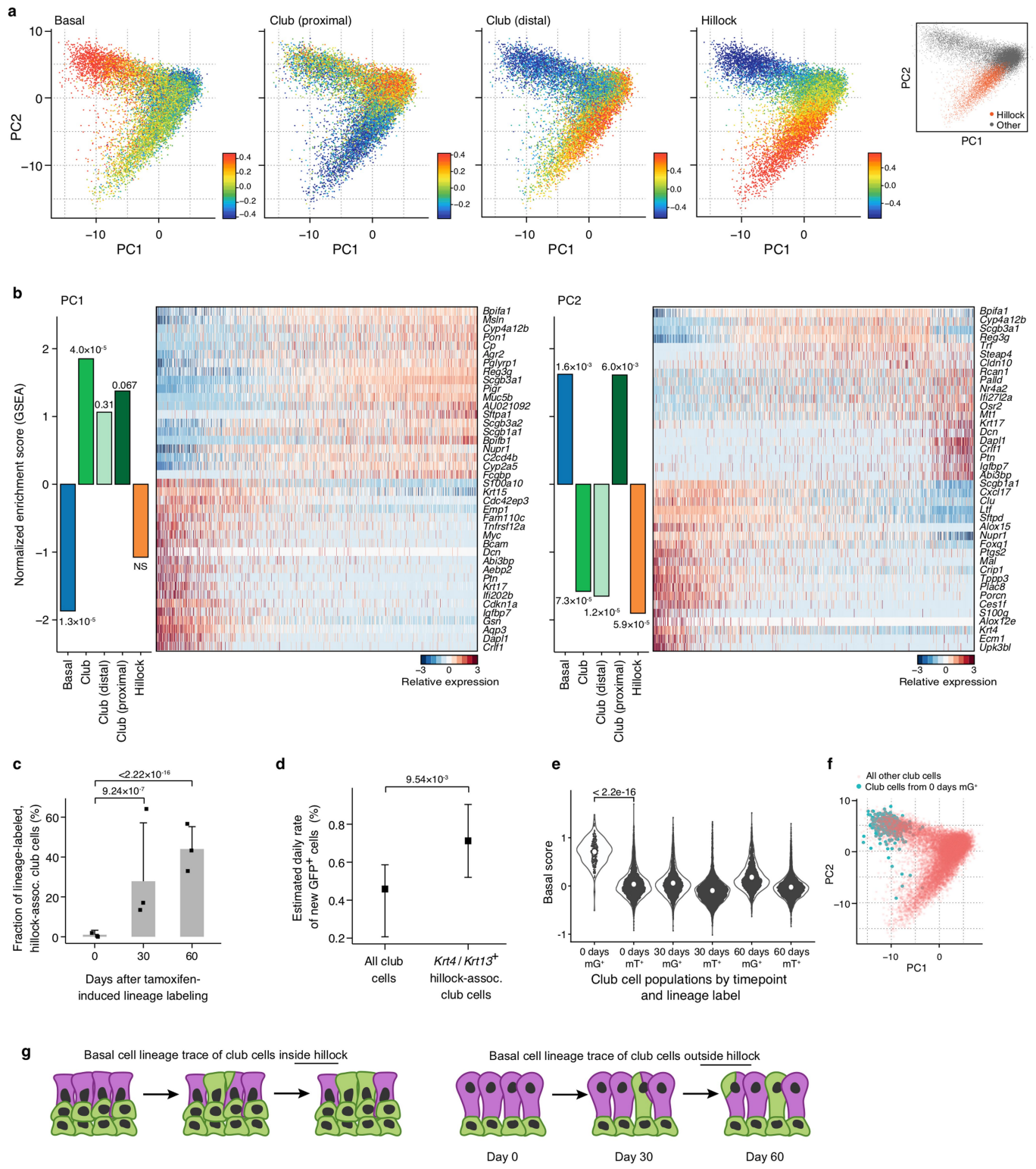
Extended Data Fig. 4 | Krt13$^+$ progenitors express a unique set of markers distinct from mature club cells. a, Proximal versus distal specific club cell expression. Relative expression level (row-wise *Z* score, colour bar) for genes (rows) enriched in proximal and distal tracheal club cells (FDR <0.05, LRT) in the full-length scRNA-seq data. **b**, Distal epithelia differentiate into mucous metaplasia. Goblet cell quantification (ln(Muc5ac$^+$/EGFP$^+$ ciliated cells)) in *Foxj1*-EGFP mice ($n = 6$, dots) in each of four conditions in (Fig. 2a). *P* values, Tukey's HSD test; black bars, mean; error bars, 95% CI. **c**, Krt8 does not distinguish pseudostratified club cell development from hillock-associated club cell development. Diffusion map embedding of 6,905 cells (as in Fig. 2b) coloured either by their Krt13$^+$ hillock membership (left, green), or by expression ($\log_2$(TPM+1)) of specific genes (all other panels). **d**, Immunostaining of hillock strata. Left: Krt13$^+$ (green) and Trp63$^+$ (magenta) basal (solid outline) and suprabasal (dashed outline) cells. Right: Krt13$^+$ (green) and Scgb1a1$^+$ (magenta, solid outline) luminal cells. Representative immunostaining from 3 mice. **e**, **f**, Krt13$^+$ hillock cells are highly proliferative. **e**, Co-stain of EdU (magenta) and Krt13 (green), representative of $n = 4$ mice. **f**, Fraction of EdU$^+$ epithelial cells in hillock (mean, 7.7%, 95% CI [4.8–10.5%]) and non-hillock (mean, 2.4%, 95% CI [1.8–3.1%]) areas. *P* values: LRT, $n = 4$ mice; black bar, mean; error bars, 95% CI. **g**, Fraction of Krt13$^+$ hillock cells that are club cell lineage labelled (%) decreases from day 5 (10.2%, 95% CI [0.07, 0.16]) to day 80 (5.2%, 95% CI [0.03, 0.08]). Error bars, 95% CI; $n = 3$ mice (dots); *P* values, LRT. **h**, Differential expression ($\log_2$(fold-change)) and associated significance ($\log_{10}$(FDR)) for each gene (dot) that is differentially expressed in *Krt13*$^+$ cells (identified using clustering in diffusion map space) compared to all cells (FDR <0.05, LRT). Colour code, cell type with highest expression (for example, green shows genes that are most highly expressed in *Krt13*$^+$ hillock cells). Dots show all the genes differentially expressed (FDR <0.05) between *Krt13*$^+$ hillock cells and other cells. Genes with $\log_2$ fold-change >1 are marked with large points, whereas others are identified as small points (grey). **i**, Enriched pathways in *Krt13*$^+$ hillock cells. Representative MSigDB gene sets (rows) that are significantly enriched (colour bar, $-\log_{10}$(FDR), hypergeometric test) in *Krt13*$^+$ hillock cells.



Extended Data Fig. 5 | Genes associated with cell fate transitions. **a**–**h**, Relative mean expression (loess-smoothed row-wise *Z* score of mean $\log_2$(TPM+1)) of significantly ($P < 0.001$, permutation test) varying genes (**a**–**d**) and transcription factors (**e**–**h**) across subsets of 6,905 (columns) basal, club and ciliated cells. Cells are pseudotemporally ordered (*x* axis, all plots) using diffusion maps (Fig. 2b, Extended Data Fig. 4c). Each cell was assigned to a cell fate transition if it was within $d < 0.1$ of an edge of the convex hull of all points (in which d is the Euclidean distance in diffusion space) assigned to that edge.

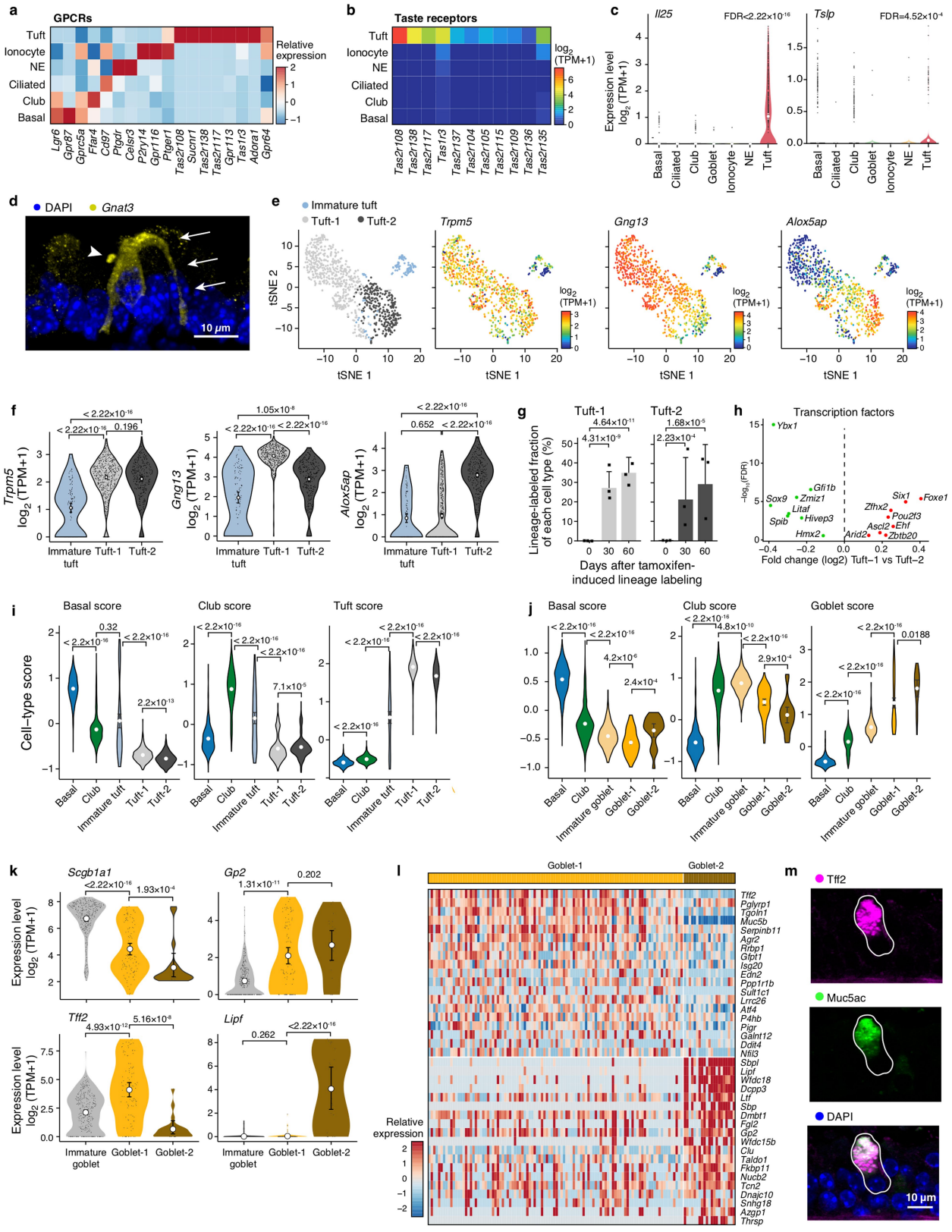


Extended Data Fig. 6 | Lineage tracing using pulse-seq. a, Schematic of the pulse-seq experimental design. **b**, Post hoc cluster annotation by known cell type markers. *t*-SNE of 66,265 scRNA-seq profiles (points) from pulse-seq, coloured by the expression ($\log_2$(TPM+1)) of single marker genes for a particular cell type or cell-cycle score (bottom right) **c**, Pulse-seq lineage-labelled fraction of various cell populations over time. Linear quantile regression fits (trendline) to the fraction of lineage-labelled cells of each type ($n = 3$ mice per time point, dots) as a function of the number of days after tamoxifen-induced labelling. β, estimated regression coefficient, interpreted as daily rate of new lineage-labelled cells; p, P value for the significance of the relationship, Wald test. As expected, goblet and ciliated cells are labelled more slowly than club cells (Fig. 3d). **d**, Labelled fraction of basal cells is unchanged during pulse-seq time course, as expected. Estimated fraction (%) of cells of each type that are positive for the fluorescent lineage label (by FACS) in each of $n = 3$ mice (points) per time point. P values, LRT; error bars, 95% CI. **e**, Proportion of basal cell lineage-labelled tuft cells at day 0 (0%; $n = 2$ mice, dots) and day 30 (22.9%, 95% CI [0.17, 0.30]; bars, estimated proportions; $n = 3$ mice). Error bars, 95% CI; P values, LRT. **f**–**h**, Conventional *Scgb1a1* (*CC10*) lineage trace of rare epithelial types shows minimal contribution to rare cell lineages. Fraction of *Scb1a1* labelled (club cell trace) cells (%) of Gnat3$^+$ tuft cells (**f**) at day 0 ($n = 3$ mice; 0.6%, 95% CI [0.00, 0.04]) and day 30 ($n = 2$ mice; 6.3%, 95% CI [0.04, 0.11]), EGFP(*Foxi1*)$^+$ ionocytes at day 30 ($n = 2$ mice; 2.9%, 95% CI [0.01, 0.11]) (**g**), and Chga$^+$ neuroendocrine cells at day 0 ($n = 2$ mice; 2.5%, 95% CI [0.01, 0.08]) and day 30 ($n = 2$ mice; 2.6%, 95% CI [0.01, 0.08]) (**h**) after club cell lineage labelling. P values, LRT; error bars, 95% CI.



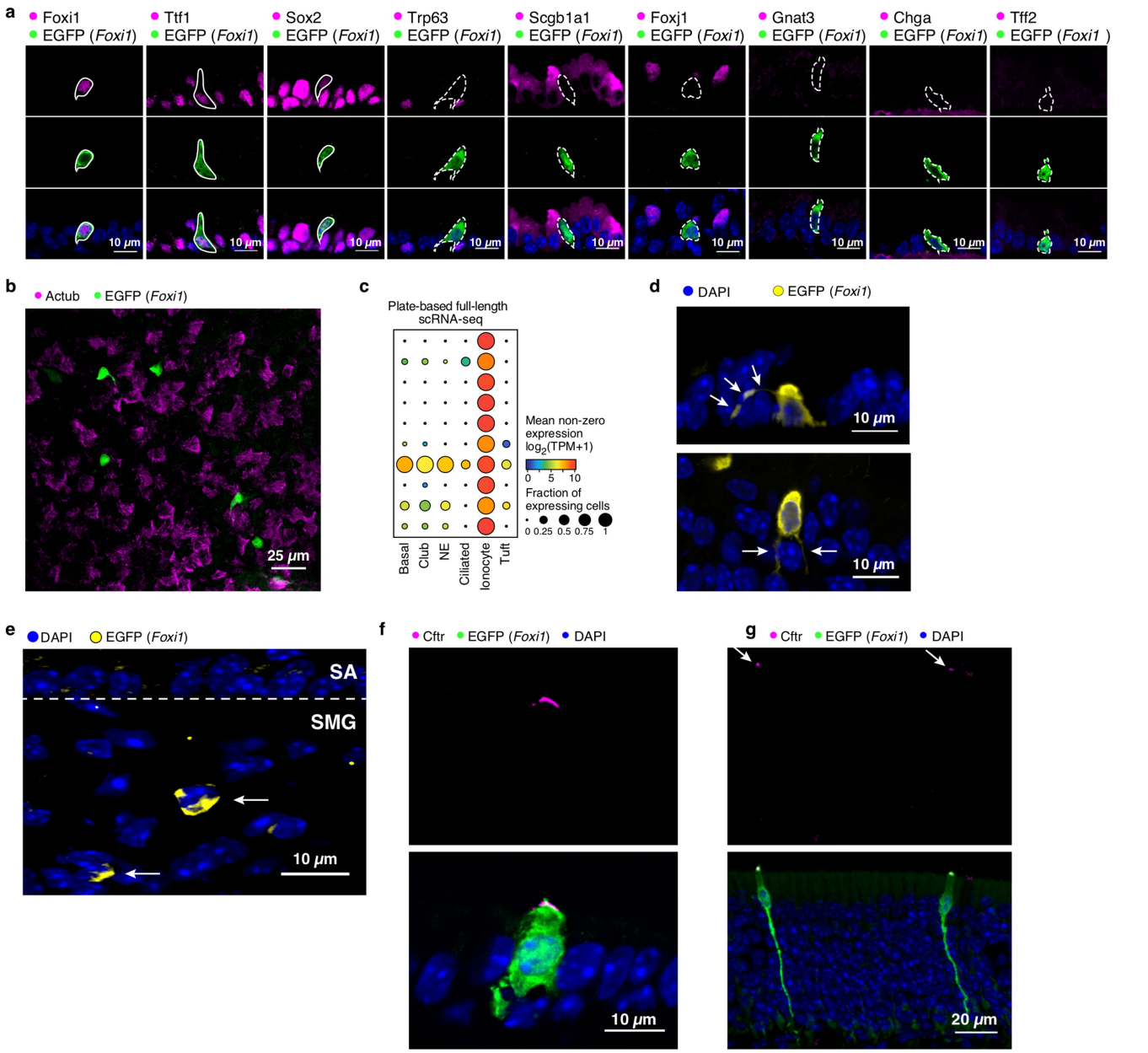
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Club cell heterogeneity and lineage tracing hillock-associated club cells using pulse-seq. a, **b**, Principal components are associated with basal to club differentiation (PC-1), proximodistal heterogeneity (PC-2), and hillock gene modules (PC-2). **a**, PC-1 (*x* axis) versus PC-2 (*y* axis) for a PCA of 17,700 scRNA-seq profiles of club cells (points) in the pulse-seq dataset, coloured by signature scores for basal (left), proximal club cells (centre left), distal club cells (centre right), the $Krt13^{+}/Krt4^{+}$ hillock (right), or their cluster assignment (inset, right). **b**, Bar plots show the extent (normalized enrichment score) and significance of association of PC-1 (left) and PC-2 (right) for gene sets associated with different airway epithelial types (*x* axis), or gene modules associated with proximodistal heterogeneity (Extended Data Fig. 4a). Heat maps show the relative expression level (row-wise *Z* score of $\log_2(\mathrm{TPM}+1)$ expression values, colour bar) of the 20 genes with the highest and lowest loadings on PC-1 (left) and PC-2 (right) in each club cell (columns, down-sampled to 1,000 cells for visualization only). *P* values, permutation test. **c**, Pulse-seq lineage tracing of hillock-associated cells. Estimated fraction (%) of cells of each type that are positive for the fluorescent lineage label (by FACS) from $n = 3$ mice (points) per time point. *P* values, LRT. Error bars, 95% CI. **d**, Hillock-associated club cells are produced at a higher rate than all club cells. Estimated rate (%) based on the slope of quantile regression fits to the fraction of lineage-labelled cells of each type. *P* values, rank test; error bars, 95% CI. **e**, **f**, Club cells initially labelled by pulse-seq are associated with basal to club cell differentiation. **e**, Distribution of basal signature scores for individual club cells (points) from each pulse-seq time point and lineage label status. *P* value, Mann–Whitney *U* test. Violin plots show the Gaussian kernel probability densities of the data, large white point shows the mean. **f**, PC-1 versus PC-2 for a PCA of 17,700 scRNA-seq profiles of club cells (points), as in **a**, highlighting club cells that are lineage-labelled at the initial time point (legend). **g**, Schematic of the more rapid turnover of basal to club cells inside (top) and outside (bottom) hillocks.

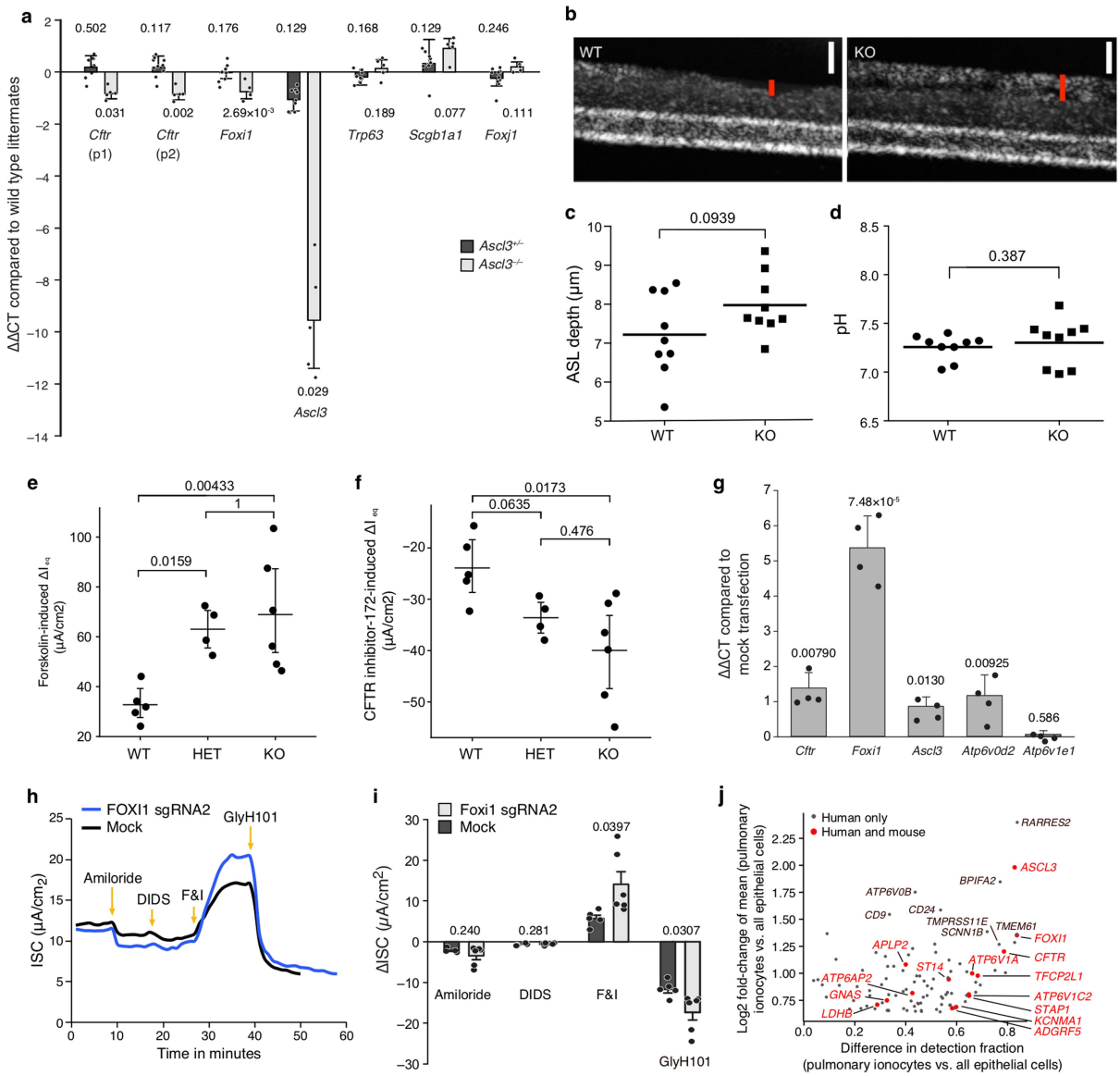


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Heterogeneity of rare tracheal epithelial cell types. a, Cell-type-enriched GPCRs. Relative expression (*Z* score of mean $\log_2$(TPM+1)) of the GPCRs that are most enriched (FDR <0.001, LRT) in the cells of each tracheal epithelial cell type based on full-length scRNA-seq data. **b**, Tuft cell-specific expression of type I and type II taste receptors. Expression level (mean $\log_2$(TPM+1)) of tuft-cell enriched (FDR <0.05, LRT) taste receptor genes in each tracheal epithelial cell type based on full-length scRNA-seq data. **c**, Tuft cell-specific expression of the type-2 immunity-associated alarmins *Il25* and *Tslp*. Expression level, of *Il25* (left) and *Tslp* (right) in each cell type. FDR, LRT. Violin plots show the Gaussian kernel probability densities of the data. **d**, Morphological features of tuft cells. Immunofluorescence staining of the tuft-cell marker Gnat3 (yellow) along with DAPI (blue). Arrowhead, 'tuft'; arrows, cytoplasmic extension. **e**, **f**, Tuft-1 and tuft-2 sub-clusters. **e**, *t*-SNE visualization of 892 tuft cells (points) coloured either by their cluster assignment (left, colour legend), or by the expression level of marker genes for mature tuft cells (*Trpm5*), tuft-1 (*Gng13*), tuft-2 (*Alox5ap*) subsets. **f**, Distribution of expression levels of the top markers for each subset. Violin plots show the Gaussian kernel probability densities of the data, large white point shows the mean. FDR, LRT, $n = 15$ mice. **g**, Tuft-1 and tuft-2 subtypes are each generated from basal cell parents. Estimated fraction of cells of each type that are positive for the basal-cell lineage label (by FACS) from $n = 3$ mice (points) per time point in the pulse-seq experiment. *P* values, LRT; error bars, 95% CI. **h**, Differential expression of tuft cell-associated transcription factors between tuft cell subtypes. Labelled genes are differently expressed in the tuft cell subsets (FDR <0.01, LRT). **i**, **j**, Mature and immature subsets are identified using marker gene expression. The distribution of expression of scores (using top 20 marker genes, Supplementary Table 1, Methods) for tuft (**i**), goblet (**j**), basal and club cells (label on top) in each cell subset (basal and club cells down-sampled to 1,000 cells). *P* values, Mann–Whitney *U* test. **k**, **l**, Gene signatures for goblet-1 and goblet-2 subsets. The distribution (**k**) and relative expression level (**l**) of marker genes that distinguish ($\log_2$ fold-change >0.1, FDR <0.001, LRT) cells in the goblet-1 and goblet-2 sub-clusters (colour bar, top and left) from the combined 3′ scRNA-seq datasets. **m**, Immunofluorescence staining of the goblet-1 marker Tff2 (magenta), the known goblet cell marker Muc5ac (green) and DAPI (blue). Solid white line: boundary of a goblet-1 cell.



Extended Data Fig. 9 | Ionocyte characterization in situ. **a**, Immunofluorescence characterization of ionocytes. Ionocytes visualized in *Foxi1*-EGFP mouse. EGFP(*Foxi1*) appropriately marks Foxi1 antibody-positive cells (left, solid outline). EGFP(*Foxi1*)$^+$ cells express canonical airway markers Ttf1 (Nkx2-1) and Sox2 (solid outlines). EGFP(*Foxi1*)$^+$ cells do not label with basal (Trp63), club (Scgb1a1), ciliated (Foxj1), tuft (Gnat3), neuroendocrine (Chga) or goblet (Tff2) cell markers (dashed outlines). **b**, Ionocytes are sparsely distributed in the surface epithelium. Representative whole-mount confocal image of ionocytes EGFP(*Foxi1*) and ciliated cells (AcTub). **c**, Expression level of ionocyte markers (rows ordered as in Fig. 5a, FDR <0.05 LRT, full-length scRNA-seq dataset) in each airway epithelial cell type. **d**, EGFP(*Foxi1*)$^+$ ionocytes extend cytoplasmic appendages (arrows). **e**–**g**, Immunofluorescence labelling of EGFP(*Foxi1*)$^+$ cells in airway regions. Submucosal gland (SMG, **e**), nasal respiratory epithelium (**f**) and olfactory neuroepithelium (**g**). Dotted line separates surface epithelium (SA) from SMG.



Extended Data Fig. 10 | Functional characterization of ionocytes. **a**, *Ascl3*(KO) moderately decreases ionocyte transcription factors and *Cftr* in ALI-cultured epithelia. Quantification ($\Delta\Delta C_T$) of expression in ionocyte (*Cftr*: −0.82 $\Delta\Delta C_T$, 95% CI [±0.20]; *Foxi1*: −0.75 $\Delta\Delta C_T$, 95% CI [±0.28]; *Ascl3*: −10.28 $\Delta\Delta C_T$, 95% CI [±1.85]) and basal (*Trp63*), club (*Scgb1a1*) or ciliated (*Foxj1*) markers in hetero- and homozygous *Ascl3* KO (colour legend) are normalized to wild-type littermates. The mean of independent probes (p1 and p2) was used for *Cftr*. $n = 10$ ($Ascl3^{+/-}$), 5 ($Ascl3^{-/-}$), 4 (wild-type) mice. *P* values: Holm–Sidak test; error bars, 95% CI. **b**, Altered ASL reflectance intensity in *Foxi1*(KO) ALI culture compared to wild type. Representative μOCT image of ASL. Red bar, airway surface liquid depth (including the periciliary and mucus layers). Scale bar (white), 10 μm. **c**, **d**, Ionocyte depletion or disruption does not affect ASL depth (**c**) as determined by μOCT, nor pH (**d**) in cultured epithelia derived from homozygous *Foxi1*(KO) ($n = 9$) versus wild type littermates ($n = 9$ mice). *P* values, Mann–Whitney *U* test. **e**, **f**, Increased ΔI_{eq} in *Foxi1*(KO) epithelia. ΔI_{eq} (*y* axis) in ALI cultures of wild type (WT), heterozygous (HET) and *Foxi1*(KO) mice ($n = 5$ (WT), $n = 4$ (HET), $n = 6$ (KO)) that were characterized for their forskolin-inducible equivalent currents (**e**; I_{eq}) and for currents sensitive to $CFTR_{inh}$-172 (**f**). The inhibitor-sensitive ΔI_{eq} values reported may underestimate the true inhibitor-sensitive current, as the inhibitor response failed to reach a steady plateau for some samples during the time scale of the experiment. **g**–**i**, *Foxi1* transcriptional activation (*Foxi1*-TA) in ferret increases *Cftr* expression and chloride transport. **g**, qRT–PCR expression quantification ($\Delta\Delta C_T$) of ionocyte markers in ferret *Foxi1*-TA ALI ($n = 4$ ferrets) normalized to mock transfection (*Cftr*: −1.39 $\Delta\Delta C_T$, 95% CI [±0.44]; *Foxi1*: −5.37 $\Delta\Delta C_T$, 95% CI [±0.91]; *Ascl3*: −0.87 $\Delta\Delta C_T$, 95% CI [±0.27]; *Atp6v0d2*: −1.18 $\Delta\Delta C_T$, 95% CI [±0.58] and *Atp6v1e1*: −0.070 $\Delta\Delta C_T$, 95% CI [±0.11]), *P* values, *t*-test; bars, means; error bar, 95% CI. **h**, **i**, *Foxi1* activation in ferret cell cultures results in a CFTR inhibitor-sensitive short-circuit current (ΔI_{sc}). Representative trace (**h**) and quantification (**i**) of short-circuit current (I_{sc}) tracings from *Foxi1*-TA ferret ALI after sgRNA reverse transfection ($n = 6$, light blue) versus mock transfection ($n = 5$, black). **j**, Evolutionarily conserved ionocyte signatures. Difference in fraction of cells in which transcript is detected and $\log_2$ fold-change between human ionocytes and all other bronchial epithelial cells. Labelled genes are differentially expressed ($\log_2$ fold-change >0.25 and FDR $<10^{-10}$, Mann–Whitney *U* test). Red, consensus ionocyte markers between mouse and human ($\log_2$ fold-change >0.25, FDR $<10^{-5}$, LRT).

ARTICLE

https://doi.org/10.1038/s41586-018-0409-3

Genetic and transcriptional evolution alters cancer cell line drug response

Uri Ben-David[1], Benjamin Siranosian[1], Gavin Ha[1,2], Helen Tang[1], Yaara Oren[1,3], Kunihiko Hinohara[1,2], Craig A. Strathdee[1], Joshua Dempster[1], Nicholas J. Lyons[1], Robert Burns[2], Anwesha Nag[2], Guillaume Kugener[1], Beth Cimini[1], Peter Tsvetkov[1], Yosef E. Maruvka[1], Ryan O'Rourke[1,2], Anthony Garrity[1], Andrew A. Tubelli[1], Pratiti Bandopadhayay[1,2,3], Aviad Tsherniak[1], Francisca Vazquez[1], Bang Wong[1], Chet Birger[1], Mahmoud Ghandi[1], Aaron R. Thorner[2], Joshua A. Bittker[1], Matthew Meyerson[1,2,3], Gad Getz[1,4], Rameen Beroukhim[1,2,3,5,7]* & Todd R. Golub[1,2,3,6,7]*

Human cancer cell lines are the workhorse of cancer research. Although cell lines are known to evolve in culture, the extent of the resultant genetic and transcriptional heterogeneity and its functional consequences remain understudied. Here we use genomic analyses of 106 human cell lines grown in two laboratories to show extensive clonal diversity. Further comprehensive genomic characterization of 27 strains of the common breast cancer cell line MCF7 uncovered rapid genetic diversification. Similar results were obtained with multiple strains of 13 additional cell lines. Notably, genetic changes were associated with differential activation of gene expression programs and marked differences in cell morphology and proliferation. Barcoding experiments showed that cell line evolution occurs as a result of positive clonal selection that is highly sensitive to culture conditions. Analyses of single-cell-derived clones demonstrated that continuous instability quickly translates into heterogeneity of the cell line. When the 27 MCF7 strains were tested against 321 anti-cancer compounds, we uncovered considerably different drug responses: at least 75% of compounds that strongly inhibited some strains were completely inactive in others. This study documents the extent, origins and consequences of genetic variation within cell lines, and provides a framework for researchers to measure such variation in efforts to support maximally reproducible cancer research.

Human cancer cell lines have facilitated fundamental discoveries in cancer biology and translational medicine[1]. An implicit assumption has been that cell lines are clonal and genetically stable, and therefore that results obtained in one study can be readily extended to another. However, findings involving cancer cell lines are often difficult to reproduce[2,3], leading investigators to conclude that the findings were either weak or the studies not carefully conducted. For example, although pharmacogenomic profiling of large collections of cancer cell lines have proven to be mostly reproducible, some discrepancies in drug sensitivity remain unexplained[4–11]. We hypothesized that cancer cell lines are neither clonal nor genetically stable, and that this instability can generate variability in drug sensitivity.

Cross-laboratory comparisons

To test the hypothesis that clonal variation exists within established cell lines, we reanalysed whole-exome sequencing data from 106 cell lines generated by both the Broad Institute (the Cancer Cell Line Encyclopedia (CCLE)) and the Sanger Institute (the Genomics of Drug Sensitivity in Cancer (GDSC)), using the same analytical pipeline for both datasets (Methods).

As expected, estimates of the allelic fraction of germline variants were nearly identical across the two datasets (median $r = 0.95$), indicating that sequencing artefacts do not substantially contribute to the erroneous appearance of low allelic fraction calls. However, the degree of agreement in allelic fraction for somatic variants was substantially lower (median $r = 0.86$; $P < 2 \times 10^{-16}$; Fig. 1a, Extended Data Fig. 1a and Supplementary Table 1). Moreover, a median of 19% of the detected non-silent mutations (range, 10–90%) were identified in only one of the two datasets (Extended Data Fig. 1b). Similarly, 26% of genes that had copy number alterations (CNAs; which are also known as copy number variants) (range, 7–99%) were discordant (Extended Data Fig. 1c–e). These results indicate that genetic variability across cultures of the same cell line is common. Indeed, a median of 22% of the genome was estimated to be affected by subclonal events across 916 CCLE cell lines (Extended Data Fig. 1f), suggesting that changes in subclonal composition may underlie the observed differences.

Genetic variation across 27 MCF7 strains

We performed extensive genomic characterization of 27 versions (hereafter called 'strains') of the commonly used oestrogen receptor (ER)-positive breast cancer cell line MCF7[12–14] (Methods, Extended Data Figs. 1g–n, 2a, b and Supplementary Table 2), including 19 strains that had not undergone drug treatment or genetic manipulation, 7 strains that carried a genetic modification generally considered to be neutral (for example, introduction of a reporter gene, Cas9 or a DNA barcode), and one strain (MCF7-M) that had been expanded in vivo in mice following anti-oestrogen therapy. Strain M was found to be an outlier, consistent with having been through strong bottlenecks, and was therefore excluded from downstream quantitative analyses.

Ten chromosome arms (25% of the genome) were differentially gained or lost in a pairwise comparison of strains (Supplementary Table 3). We detected 283 genes with copy number gains and 405 genes with copy number losses (compared to basal ploidy) in at least one strain. Only a small minority of these changes (13% of gains and 21% of losses) were detected in all strains. Of these changes, 7% of gains and 13% of losses were detected in only a single strain, and the remaining events were observed variably across strains (Fig. 1b and

[1]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Dana-Farber Cancer Institute, Boston, MA, USA. [3]Harvard Medical School, Boston, MA, USA. [4]Massachusetts General Hospital, Boston, MA, USA. [5]Brigham and Women's Hospital, Boston, MA, USA. [6]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [7]These authors jointly supervised this work: Rameen Beroukhim, Todd R. Golub. *e-mail: rameen_beroukhim@dfci.harvard.edu; golub@broadinstitute.org

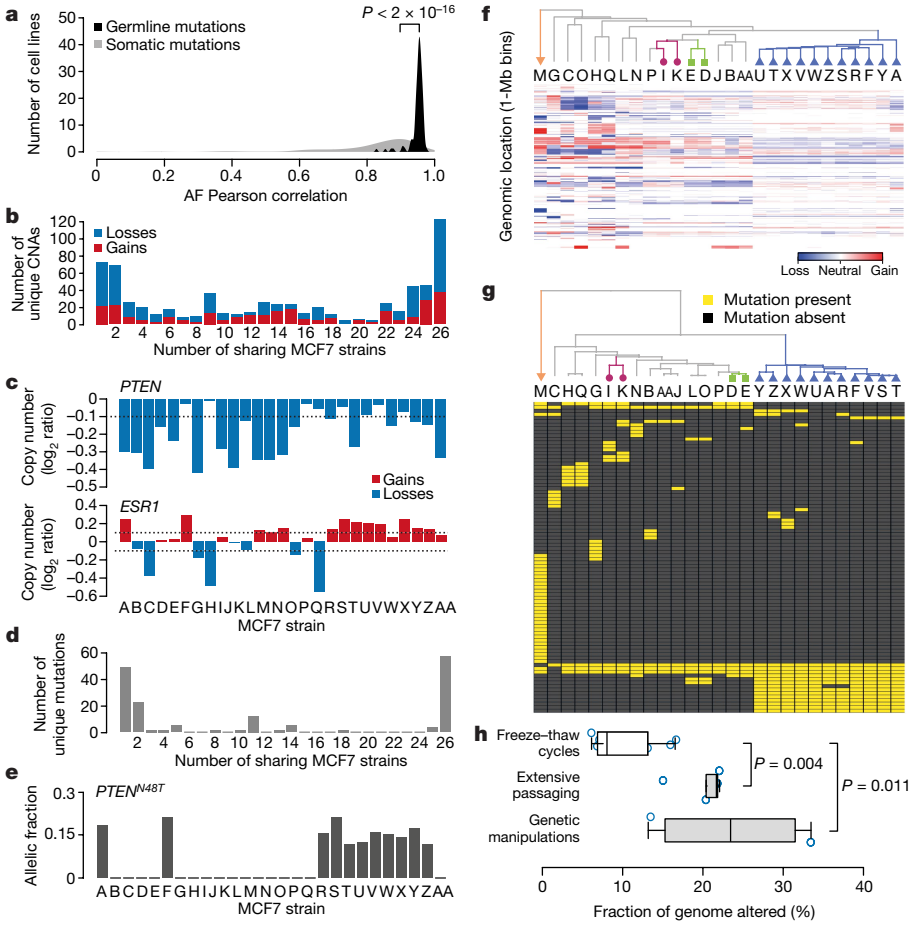


Fig. 1 | Extensive genetic variation across 27 strains of the cancer cell line MCF7. a, The distribution of pairwise allelic fraction (AF) correlations between the Broad and the Sanger cell lines ($n = 106$), for germline (black) and somatic (grey) SNVs. One-tailed paired Wilcoxon rank-sum test. **b**, The number of gene-level CNAs shared by each number of MCF7 strains. Red, gains; blue, losses. **c**, CNAs of two genes, *PTEN* and *ESR1*. **d**, The number of non-silent point mutations shared by each number of MCF7 strains. **e**, The allelic fraction of inactivating mutations in the tumour suppressor *PTEN*. **f**, Top, unsupervised hierarchical clustering of 27 MCF7 strains based on CNA profiles derived from low-pass whole-genome sequencing. Orange, strain M subjected to in vivo passaging and drug treatment; blue, 11 connectivity map strains cultured in the same laboratory without extensive passaging; green, strains D and E cultured in the same laboratory and separated by few passages; purple, strains I and K separated by Cas9 introduction. Bottom, corresponding heat map of the CNA landscapes of the strains relative to the median CNA landscape. Red, gains; blue, losses. **g**, Top, unsupervised hierarchical clustering of 27 MCF7 strains, based on their non-silent SNV profiles derived from deep targeted sequencing. Colours as in **f**. Bottom, corresponding heat map of the mutation status of non-silent mutations across strains. Mutations that were identified in a subset of the strains at AF > 0.05 are shown. Yellow, mutation present; grey, mutation absent. **h**, Comparison of the magnitude of CNAs observed following multiple freeze–thaw cycles ($n = 9$; R, A and S versus W, X and Y), extensive passaging ($n = 5$; D versus L versus AA, B versus I and P), and genetic manipulations ($n = 4$; AA versus O, B versus C, I versus J and K). Bar, median; box, 25th and 75th percentiles; whiskers, 1.5× IQR of lower and upper quartile; circles, data points. Two-tailed Wilcoxon rank-sum test.

Supplementary Table 4). The differential events included genes commonly gained or lost in breast cancer (for example, *TP53*, *PTEN*, *EGFR*, *PIK3CA* and *MAP2K4*; Extended Data Fig. 3a). For example, *PTEN* was deleted in 17 strains and retained in the other 10 (Fig. 1c). Similarly, the oestrogen receptor gene *ESR1* was gained in 12 strains, lost in 6 and unaltered in 9 (Fig. 1c), and this correlated with differential expression of ERα ($P = 0.009$; Extended Data Fig. 3b, c and Supplementary Discussion).

Genetic variation was similarly observed at the level of point mutations, small insertions or deletions (indels) and chromosomal translocations. Only 35% of 95 non-synonymous single nucleotide variants (SNVs) and indels that affected the coding sequence or splice regions were shared by all strains: 29% were unique to a single strain, and the remaining were present in a subset of strains (Fig. 1d, e, Extended Data Fig. 3dj, Supplementary Tables 5, 6 and Supplementary Discussion). Similar, albeit lower, variability was observed among mutations listed as recurrent in the COSMIC database[15], consistent with COSMIC mutations tending to be clonal mutations of the founding populations (Extended Data Fig. 3f).

Unsupervised hierarchical clustering analysis, in which genetic distance was reflected by the branch lengths of the dendrogram, generated a branch structure that accurately reflected the history of the strains. For example, strain M, which had been subjected to in vivo passaging and drug treatment, was the most genetically distinct; the 11 strains used by the connectivity map project[16] over a 10-year period clustered tightly together; and sibling strains D and E, which were only a few passages apart, were the closest to each other (Fig. 1f, g and Extended Data Fig. 3g). The genetic distance between strains appeared to be affected more by passage number and genetic manipulation than by freeze–thaw cycles (Fig. 1h and Extended Data Fig. 4).

Sources of variation

Analysis of variant allelic fractions revealed extensive subclonality across strains (Fig. 2a, b and Extended Data Fig. 5a). For example, all 27 strains had a *PIK3CA*-activating mutation (G1633A), but the allelic fraction varied from 0.21 to 0.70 (Extended Data Fig. 5b). On the basis of allelic fractions and copy number status, 45% of all observed mutations were determined to be subclonal ($P < 0.01$ in a

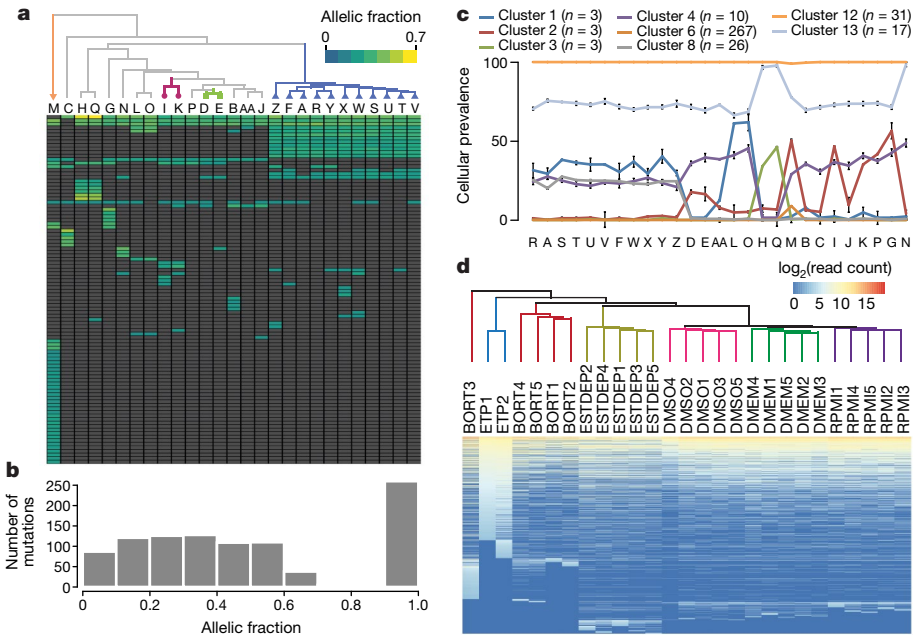


Fig. 2 | Genetic heterogeneity and clonal dynamics underlying genetic variation. a, Top, unsupervised hierarchical clustering of 27 MCF7 strains based on the allelic fractions of their non-silent SNVs. Colours as in Fig. 1. Bottom, corresponding heat map of the allelic fractions of non-silent mutations present in a subset of the strains. **b**, The distribution of allelic fractions of non-silent mutations across strains. **c**, The cellular prevalence of mutation clusters across MCF7 strains identified by a PyClone analysis. Mutation clusters with differential abundance (a difference in cellular prevalence (ΔCP) > 0.15), the clonal cluster (cluster 6; CP ≈ 1 in all clones) and a cluster unique to MCF7-M (cluster 12) are shown. n = mutations per cluster, data are mean ± s.e.m. **d**, Top, unsupervised hierarchical clustering of 27 samples of DNA-barcoded MCF7-D based on barcode representation. Dendrogram branches are coloured by culture condition. Bottom, corresponding heat map of barcode representation. ETP, early time point; RPMI, RPMI 1640 medium; DMEM, DMEM medium; DMSO, RPMI 1640 with 0.05% DMSO; ESTDEP, oestrogen-depleted RPMI 1640 medium; BORT, bortezomib (500 nM; 48 h exposure) followed by RPMI 1640.

binomial test). PyClone[17,18], which reconstructs subclonal structure by clustering mutations with similar cellular prevalence, found multiple subclones within each MCF7 strain, with varying abundance across strains (Fig. 2c). Indeed, for 43% of the non-silent SNVs, cellular prevalence differed by >50% across strains (Extended Data Fig. 5c, d and Supplementary Table 7).

We next investigated whether clonal dynamics were stochastic or the product of selection. We barcoded MCF7 cells (strain D) and evaluated the change in barcode representation over time under five culture conditions, each in five replicates. We reasoned that if clonal dynamics were stochastic, distinct barcoded populations would emerge in independent replicates. By contrast, if pre-existing subclones were selected under different conditions, enrichment of the same barcodes would be observed in replicate cultures[19]. Unsupervised hierarchical clustering by barcode representation revealed that biological replicates clustered together (Fig. 2d and Supplementary Table 8), indicating that pre-existing subclones are indeed selected by changes in culture conditions.

Next, we characterized the genetic stability of three wild-type single-cell-derived MCF7 clones and five single-cell-derived clones with a 'neutral' genetic manipulation (stable expression of a luciferase reporter; Methods, Extended Data Fig. 5e and Supplementary Tables 9, 10). Clones derived from the same parental population differed in their mutational landscapes: a median of 15% of the non-silent SNVs detected in the wild-type parental population (range, 13% to 16%), were not observed in their single-cell-derived progeny or vice versa (Extended Data Fig. 5f, g).

Moreover, the single-cell clones continued to evolve into heterogeneous populations. We propagated two clones for 8–14 months and sequenced their DNA at multiple time points (Supplementary Tables 9, 10). A median of 13% of the non-silent SNVs (range, 8–16%) were not shared between time points (Extended Data Fig. 5g). Similar results were observed based on cytogenetic analysis (Extended Data Fig. 5h–k and Supplementary Table 11), indicating that even single-cell-derived clones are genomically unstable.

Gene expression variation

We next measured transcriptomic variation across the MCF7 strains using the L1000 assay[16,20,21] (Supplementary Table 12). Despite an overall similarity in their global gene expression profiles (Fig. 3a and Extended Data Fig. 6a), the 27 strains also showed extensive expression variation: 654 genes (median; range, 10–1,574) were differentially expressed by at least twofold between pairs of strains ($P < 0.05$, $Q < 0.05$), and the differentially expressed genes converged on important biological pathways (Extended Data Fig. 6b–d and Supplementary Table 13). Notably, the 27 strains clustered similarly in the space of mutations and expression profiles, and the expected downstream consequences of genetic mutations were observed in the gene expression variation (Figs. 1f, g, 3b–g, Extended Data Fig. 6e–i and Supplementary Table 14). For example, strains with inactivating *PTEN* mutations or activating *PIK3CA* mutations had decreased *PTEN* and increased mTOR gene expression signatures, respectively (Fig. 3e, f and Extended Data Fig. 6g–i). Similarly, copy number loss of *ESR1* was associated with reduced oestrogen signalling (Fig. 3g and Extended Data Fig. 6g).

We further explored gene expression heterogeneity using single-cell RNA sequencing of 26,465 individual cells from two parental and four single-cell-derived clones (Methods, Extended Data Fig. 6j–r and Supplementary Discussion). Unsupervised clustering showed that cells from the single-cell-derived clones did not cluster independently, but were mixed with the parental population, indicating high similarity in overall gene expression (Fig. 3h and Extended Data Fig. 6o). Notably, the extent of expression heterogeneity among the single-cell-derived clones was not substantially lower than the heterogeneity of the parental population (Extended Data Fig. 6p), and increased with time in culture (Extended Data Fig. 6q–r, Supplementary Table 15 and Supplementary Discussion). These results indicate that variation in gene expression arises de novo, in addition to reflecting selection of pre-existing subclones[22].

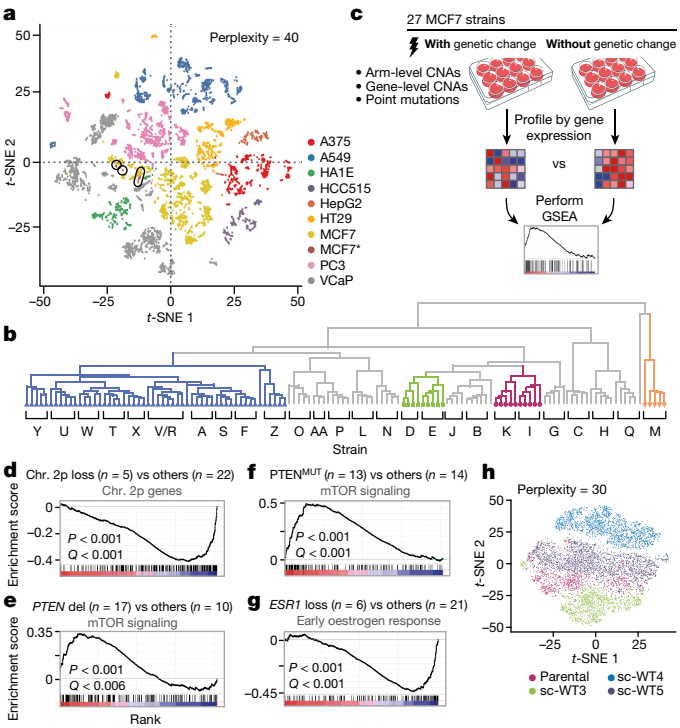


Fig. 3 | Extensive transcriptomic variation associated with genetic variation. a, A *t*-distributed stochastic neighbour embedding (*t*-SNE) plot of gene expression profiles from multiple samples of nine cancer cell lines. The 27 MCF7 strains profiled in the current study are indicated by an asterisk in the key and are encircled in the plot. **b**, Unsupervised hierarchical clustering of the strains, based on their global gene expression profiles. Colours as in Fig. 1. **c**, Schematics of the analysis performed to evaluate the association between genetic variation and transcriptional programs. **d**, Arm-level gains and losses are associated with significant up- and downregulation of genes transcribed from the aberrant arms. **e**, Gene-level CNAs are associated with significant dysregulation of the perturbed pathways. For example, upregulation of mTOR signalling was found in strains that had lost a copy of *PTEN*. **f**, Point mutations are associated with significant dysregulation of the perturbed pathways. For example, upregulation of mTOR signalling was found in strains with an inactivating *PTEN* mutation. **g**, Copy number loss of *ESR1* is associated with significant downregulation of the oestrogen response. **h**, A *t*-SNE plot of single-cell RNA-sequencing data from a parental population and three of its single-cell-derived clones. scWT3–5, single-cell wild-type clone 3–5.

Verification in additional cell lines

To exclude the possibility that the variation that we observed across MCF7 strains was unique to that cell line, we repeated genomic analyses on 23 strains of the commonly used lung cancer cell line A549[23] (Extended Data Fig. 2c, d and Supplementary Tables 16–20). We observed a similar level of molecular variation across these strains (Extended Data Fig. 7). For example, loss of *CDKN2A*, the most significantly deleted gene in lung adenocarcinomas[24], was detected in 5 strains, but a normal copy number was retained in the other 18 (Extended Data Fig. 7f). Whereas transcriptome analyses showed that oestrogen signalling was the most variable pathway in MCF7 cells (Extended Data Fig. 6c and Supplementary Table 13), *KRAS* signalling was the most variable pathway in A549 (Extended Data Fig. 7n and Supplementary Table 20), a commonly used model of *KRAS*-dependent cancer.

The generalizability of our findings was further confirmed by deep targeted sequencing of multiple strains from 11 additional cell lines (Extended Data Fig. 8 and Supplementary Tables 21–24). Notably, genomic instability was not limited to transformed cancer cell lines (Supplementary Discussion). For example, the variation across 15 strains of MCF10A[25], a non-transformed human mammary cell line, was as high as the variation that we found in MCF7 cancer cells (median discordance, 26%; range, 17–40%; Extended Data Fig. 8a, h).

Functional consequences of genomic variation

The extensive genomic variation across strains was associated with variation in biologically meaningful cellular properties. We examined several measures of basic cellular function, including doubling time and cell morphology, using quantitative live cell imaging[26] (Methods). MCF7 strains varied in doubling times by as much as 3.5-fold (median, 31 h; range, 22–78 h; Extended Data Fig. 9a, b). Similarly, cell size and shape were highly variable across strains (Extended Data Fig. 9c–f and Supplementary Table 25). Clustering based on morphological traits was similar to clustering based on genomics or transcriptomics (Extended Data Fig. 9g), and genomic features correlated with proliferation (Extended Data Fig. 9h, i and Supplementary Discussion).

Genomic instability also had major effect on drug response. We measured cell viability following treatment with 321 drugs at a single concentration (5 μM) across the 27 MCF7 strains (Supplementary Table 26). Of these, 55 compounds had strong activity (>50% growth inhibition) against at least one strain. However, at least one strain was entirely resistant (<20% growth inhibition) to 48 out of these 55 (87%) active compounds (Fig. 4a, b and Extended Data Fig. 10a). The same phenomenon was observed at a more stringent threshold: of 42 compounds with strong activity in at least two strains, 33 (79%) were inactive in at least two strains (Extended Data Fig. 10b–d, j and Supplementary Discussion). All 33 differentially active compounds were validated in an eight-point dose–response analysis of each of the 27 strains (median Spearman's $\rho = 0.42$ between screens, $P = 3 \times 10^{-9}$; Extended Data Fig. 10k, Supplementary Table 27 and Supplementary Discussion).

The high degree of variability in drug response cannot be explained by irreproducibility of the assay. First, replicate treatments yielded highly concordant results (median Pearson's $r = 0.97$, $P < 2 \times 10^{-16}$; Extended Data Fig. 10l). Second, compounds with the same mechanism of action had similar patterns of activity across strains (Fig. 4a, c; $P = 3 \times 10^{-7}$). For example, the same activity pattern was observed for three proteasome inhibitors (bortezomib, MG-132 and carfilzomib; Fig. 4d), and was associated with biochemically measured differential proteasome activity (Extended Data Fig. 10m–o). Third, for 82% of differentially active compounds, we found differential gene expression signatures of the mechanism of action[27] of the compounds between sensitive and insensitive strains ($P = 2 \times 10^{-5}$; Fig. 4e–h, Extended Data Fig. 10p–u and Supplementary Tables 28, 29).

Indeed, drug response was associated with transcriptional differences in relevant pathways. For example, strains sensitive to CDK inhibitors had an upregulated cell cycle signature and strains sensitive to PI3K inhibitors had an upregulated mTOR signature (Fig. 4f, g and Extended Data Fig. 10p, q). Notably, the strains that were the most resistant to treatment in general (strains M and Q) showed downregulation of drug metabolism pathways (Extended Data Fig. 10v). Differences in proliferation rate did not explain the majority of the observed differential drug activity (median Spearman's $\rho = 0.017$, $P = 0.60$; Supplementary Table 30).

Genetic variation could be linked directly to differential drug response. For example, genetic inactivation of *PTEN* was associated with decreased *PTEN* and increased *AKT* expression signatures (Figs. 1c, e, 3e, f), and increased sensitivity to the AKT inhibitor IV (Fig. 4h, i). Similarly, *ESR1* loss was associated with reduced oestrogen signalling (Figs. 1c, 3g), which was in turn associated with reduced sensitivity to tamoxifen or oestrogen depletion (Fig. 4j and Extended Data Fig. 10w–x). More broadly, clustering of the MCF7 strains based on their drug response was highly similar to clustering based on genetics or gene expression (Figs. 1g, 2a, 3b, 4a, Extended Data Fig. 11a and Supplementary Discussion). Genome-wide CRISPR screens showed that genetic dependencies were affected by genomic variation similar to pharmacological dependencies (Extended Data Fig. 11b–f, Supplementary Table 31 and Supplementary Discussion), and functional analyses revealed that single-cell-derived clones remained

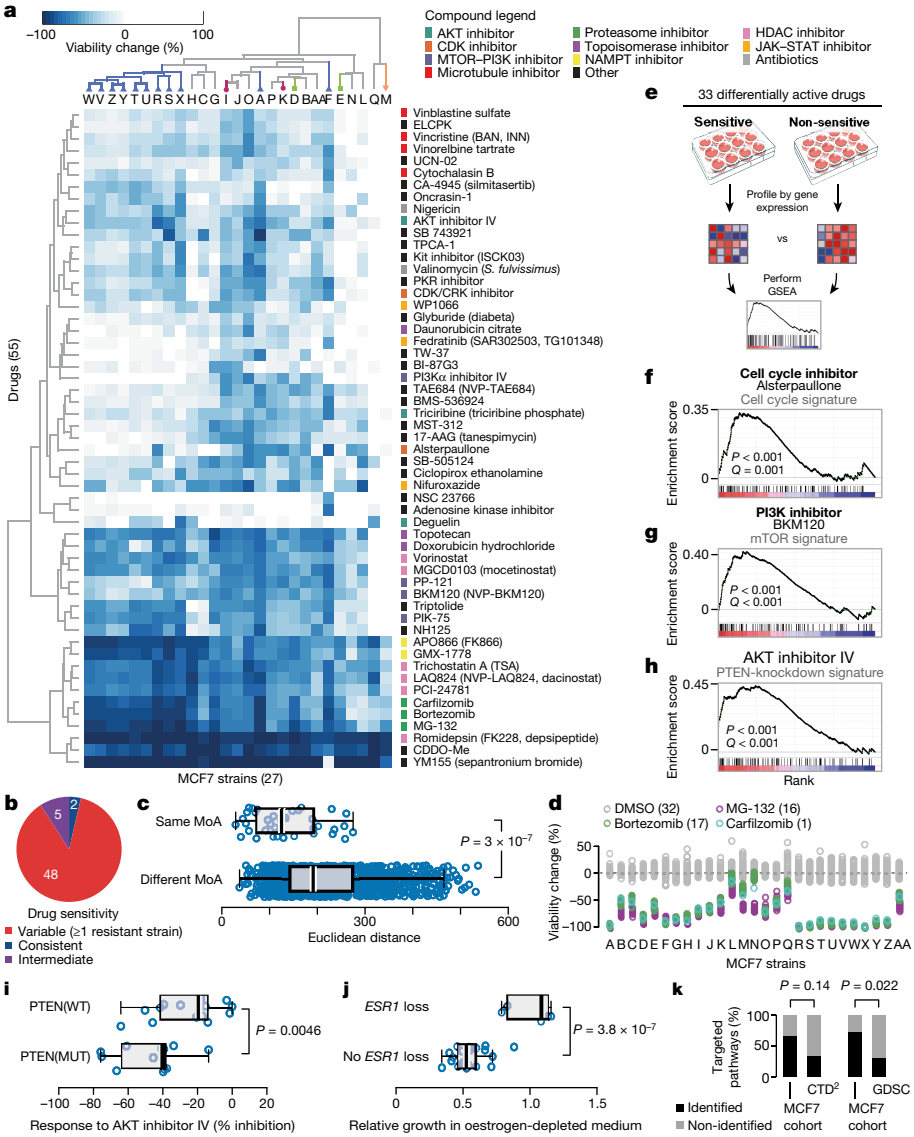


Fig. 4 | Drug-response consequences of genetic and transcriptomic variation. a, Top, unsupervised hierarchical clustering of 27 MCF7 strains based on their response to the 55 active compounds in the primary screen. Colours as in Fig. 1. Bottom, corresponding heat map of the percentage of viability change for each compound across strains. Compounds are coloured based on their mechanism of action. **b**, Classification of the screened compounds based on their differential activity. Consistent, viability change $<-50\%$ for all strains; variable, viability change $<-50\%$ for some strains and $>-20\%$ for other strains; intermediate, viability change in between these values. **c**, Comparison of the similarity in drug response patterns between compounds that share the same mechanism of action ($n=39$) and compounds that work through different mechanisms ($n=1{,}439$). One-tailed Wilcoxon rank-sum test. **d**, Highly similar differential drug response patterns for three proteasome inhibitors: bortezomib, MG-132 and carfilzomib. Each data point represents the mean of two replicates. The number of data points per strain is mentioned in parentheses. The response pattern with no drug (DMSO control) is presented for comparison. **e**, Schematics of the analysis performed to evaluate the association between drug response and transcriptional variation. **f**, Upregulation of the KEGG cell cycle signature in strains sensitive to the cell cycle inhibitor alsterpaullone (8 sensitive and 15 resistant strains). **g**, Upregulation of mTOR signalling in strains sensitive to the PI3K inhibitor BKM-120 (8 sensitive and 5 resistant strains). **h**, Upregulation of the genes that are upregulated when *PTEN* is knocked down in strains sensitive to AKT inhibitor IV (6 sensitive and 9 resistant strains). **i**, Strains with *PTEN* mutation ($n=12$) respond more strongly to AKT inhibitor IV than strains without the mutation ($n=14$). **j**, Strains with *ESR1* copy number loss ($n=5$) grow better in oestrogen-depleted medium than strains without *ESR1* loss ($n=21$). **k**, Comparison of gene set enrichment analysis-based MoA identification between the MCF7 cohort and the CTD^2 ($n=15$) and GDSC ($n=19$) cohorts across matched drugs. Two-tailed Fisher's exact test. For all box plots: bar, median; box, 25th and 75th percentiles; whiskers, 1.5× the interquartile range of the lower and upper quartile; circles, data points.

phenotypically unstable (Extended Data Fig. 11g–i and Supplementary Discussion).

We thus hypothesized that variation across otherwise isogenic strains might be harnessed to discover mechanisms of drug sensitivity and resistance. Indeed, we found that basal gene expression profiles across the 27 MCF7 strains could be more readily connected to the mechanism of action of active drugs than did larger panels of breast cancer cell lines derived from different patients[5,8] (Fig. 4k, Supplementary Table 32 and Supplementary Discussion).

Discussion

Our results show that established cancer cell lines, generally thought to be clonal, are in fact highly genetically heterogeneous. This heterogeneity results both from clonal dynamics (that is, changes in the abundance of pre-existing subclones) and from continuous instability (that is, the appearance of new genetic variants). Moreover, genetic heterogeneity leads to varying patterns of gene expression, which in turn result in differential drug sensitivity. These findings have a number of important implications, which are summarized in Extended Data Table 1.

We found that changes in clonal composition underlie much of the observed variability in cell line behaviour. Such clonal composition changes follow selection by particular conditions (for example, growth medium) or by genetic manipulations associated with a population bottleneck. The genetic distance between cell line strains was strongly correlated with their gene expression distance and with their drug-response distance. Cell line diversification can therefore be estimated using inexpensive profiling methods (Extended Data Fig. 11j). To facilitate routine assessment of cell line diversification, we have created the Cell STRAINER (strain instability profiler) portal (https://cellstrainer.broadinstitute.org), where users can upload cell line genomic data and measure their strain's genetic distance from a reference.

Variation within cancer cell lines can also be useful in at least two ways. First, deeper characterization (for example, by single-cell sequencing) of the heterogeneity within cultures of common cell lines could enable the study of cooperative and competitive interactions between cancer cell populations[28,29] and mechanisms of pre-existing drug resistance[19]. Second, owing to their matched genetic background, naturally occurring 'isogenic-like' strains could help to uncover the association between molecular features and phenotypes such as drug response.

We conclude that cancer cell lines remain a powerful tool for cancer research, but their genomic evolution leads to a high degree of variation across cell line strains, which must be considered in experimental design and data interpretation.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0409-3.

Received: 29 September 2017; Accepted: 22 June 2018;
Published online 8 August 2018.

1. Sharma, S. V., Haber, D. A. & Settleman, J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat. Rev. Cancer* **10**, 241–253 (2010).
2. Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. The economics of reproducibility in preclinical research. *PLoS Biol.* **13**, e1002165 (2015).
3. Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* **10**, 712 (2011).
4. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
5. Garnett, M. J. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
6. Basu, A. et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
7. Yang, W. et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
8. Seashore-Ludlow, B. et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.* **5**, 1210–1223 (2015).
9. Haibe-Kains, B. et al. Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
10. The Cancer Cell Line Encyclopedia & Genomics of Drug Sensitivity in Cancer Investigators. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **528**, 84–87 (2015).
11. Haverty, P. M. et al. Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature* **533**, 333–337 (2016).
12. Soule, H. D., Vazguez, J., Long, A., Albert, S. & Brennan, M. A human cell line from a pleural effusion derived from a breast carcinoma. *J. Natl Cancer Inst.* **51**, 1409–1416 (1973).
13. Brooks, S. C., Locke, E. R. & Soule, H. D. Estrogen receptor in a human cell line (MCF-7) from breast carcinoma. *J. Biol. Chem.* **248**, 6251–6253 (1973).
14. Lee, A. V., Oesterreich, S. & Davidson, N. E. MCF-7 cells–changing the course of breast cancer research and care for 45 years. *J. Natl Cancer Inst.* **107**, djv073 (2015).
15. Bamford, S. et al. The COSMIC (catalogue of somatic mutations in cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
16. Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452 (2017).
17. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
18. Eirew, P. et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
19. Bhang, H. E. et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).
20. Berger, A. H. et al. High-throughput phenotyping of lung cancer somatic mutations. *Cancer Cell* **30**, 214–228 (2016).
21. Peck, D. et al. A method for high-throughput gene expression signature analysis. *Genome Biol.* **7**, R61 (2006).
22. Gupta, P. B. et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
23. Lieber, M., Smith, B., Szakal, A., Nelson-Rees, W. & Todaro, G. A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells. *Int. J. Cancer* **17**, 62–70 (1976).
24. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
25. Soule, H. D. et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* **50**, 6075–6086 (1990).
26. Bray, M. A. et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).
27. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
28. Janiszewska, M. et al. *In situ* single-cell analysis identifies heterogeneity for *PIK3CA* mutation and *HER2* amplification in *HER2*-positive breast cancer. *Nat. Genet.* **47**, 1212–1219 (2015).
29. Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).

Acknowledgements We thank G. Wei, W. Zhang, C. Mader, J. Roth, D. Thomas, A. Kung, D. Davison, C. Chouinard, K. Walsh, A. Navarro, A. Berger, D. Wheeler, X. Jin, A. Hong, M. Trakala, P. Cho, J. Kuiken, R. Boidot, X. Lu, F. Huang, C. Johannessen, X. Wu, S. Santaguida, N. Sethi, A. Amon, K. Polyak, J. Brugge, D. Yu, J. Klefstrom, W. Hahn, I. Dunn and Y. Mei for contributing cell lines for this study; M. Ducar and S. Drinan for assistance with the OncoPanel assay; Z. Herbert for assistance with the whole-genome sequencing; J. Davis, S. Johnson, D. Lahr, J. Gould, M. Macaluso, X. Lu and T. Natoli for assistance with the L1000 assay; D. Feldman for assistance with the barcoding experiment; J. McFarland, J. Shih, C. Oh, A. Cherniack and P. Clemons for computational assistance; V. Jones and J. Gale for assistance with drug screening; K. Hartland for assistance with Cell Painting staining and imaging; A. Regev, O. Rozenblatt-Rosen, A. Neumann, D. Dionne and L. Nguyen for assistance with 10X single-cell RNA sequencing; E. Gonzalez for assistance with cytogenetic analyses; L. Lichtenstein, D. Benjamin, S. K. Lee, V. Ruano-Rubio and A. Chevalier for the GATK CNA pipeline; Z. Tothova, J. Boehm, O. Cohen, C. Johannessen, A. Subramanian, A. Carpenter, I. Tirosh, Y. Brody, Y. Zeira and R. Pistofidis for discussions. U.B.-D. is supported by a HFSP postdoctoral fellowship. This work was supported by HHMI (T.R.G.), NIH (R01 CA188228; R.Be.), Gray Matters Brain Cancer Foundation (R.Be.), Bridge Project (P.B. and R.Be.), Broad Institute SPARC award (P.B. and R.Be.) and BroadNext10 trainee grant (U.B.-D.).

Reviewer information *Nature* thanks J. Yang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions U.B.-D. conceived the project, collected the data, performed the experiments and carried out the analyses; B.S. assisted with computational analyses and figure preparation; G.H. assisted with ichorCNA, PyClone analyses and building the Cell STRAINER portal; H.T. assisted with cell culture and experiments; N.J.L. assisted with the L1000 assay; R.Bu., A.N. and A.R.T. assisted with the OncoPanel assay and analysis; B.C. assisted with the Cell Painting analysis; J.A.B. assisted with the chemical screens and their analysis; P.T. assisted with the proteasome activity assay and with the western blots; P.B., R.O. and A.G. assisted with the DNA barcoding experiment; C.A.S. and K.H. derived MCF7 single-cell clones; Y.O. assisted with single-cell RNA-sequencing data analysis; J.D. and F.V. assisted with the analysis of CRISPR screens; M.G., G.K. and A.T. assisted with the comparison of the Broad and Sanger whole-exome sequencing data; Y.E.M. assisted with compiling the RPE1 RNA-sequencing dataset; C.B. and G.G. assisted with building a web portal; B.W. and A.A.T. assisted with figure design and preparation; R.Be. and T.R.G. directed the project. U.B.-D., G.H., M.M., R.Be. and T.R.G. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0409-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0409-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to R.Be. or T.R.G.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture. The MCF7, HT29, MDAM453 and A375 cell lines were cultured in RPMI 1640 (Life Technologies), with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin–streptomycin–glutamine (Life Technologies). The A549, VCaP, PC3, HCC515, HepG2, HeLa and Ben-Men-1 cell lines were cultured in DMEM (Life Technologies), with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin–streptomycin–glutamine (Life Technologies). The HA1E cell line was cultured in MEMα (Life Technologies), with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin–streptomycin–glutamine (Life Technologies). The MCF10A cell line was cultured in MEGM Mammary Epithelial Cell Growth Medium (Lonza) supplemented with the MEGM Bulletkit (Lonza). The single-cell-derived clones scWT3, scWT4 and scWT5, as well as their parental MCF7 population, were cultured in DMEM (Life Technologies), with 10% fetal bovine serum (Sigma-Aldrich), 1% penicillin–streptomycin–glutamine (Life Technologies) and 10 μg ml^{-1} insulin (Sigma-Aldrich). Cells were incubated at 37 °C, 5% CO_2 and passaged twice a week using Trypsin-EDTA (0.25%) (Life Technologies). All strains of the same cell line were cultured under the same conditions, cell identity was confirmed and the cells were confirmed to be mycoplasma-free. Cells were tested for mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza), according to the manufacturer's instructions. Cell line identity was confirmed using SNP-based DNA fingerprinting (see below).

Derivation of single-cell clones. The wild-type single-cell-derived MCF7 clones were generated by cell sorting. Single cells were sorted into individual wells of 96-well plates, using a BD FACSAriaII SORP Cell Sorter. Three resultant clones were expanded for a period of approximately three months before the experiments. The genetically manipulated single-cell-derived MCF7 *GREB1* and MCF7 *ESR1* clones were generated using CRISPR–Cas9-mediated genome engineering to insert a NanoLuciferase reporter gene into the 3′ UTR of the respective genes. In brief, a selectable reporter gene cassette was engineered using the EMCV IRES element to drive expression of the destabilized NLucP reporter gene (Promega) fused to the N terminus of the *BSR* blasticidin-resistance gene (Invivogen) containing a P2A self-cleaving peptide element. For targeting *GREB1*, the reporter gene cassette was subcloned into a construct containing approximately 2 kb of the *GREB1* gene surrounding the termination codon in exon 33, such that the reporter gene cassette is located 9 bp downstream of the *GREB1* termination codon in the resulting mRNA hybrid transcript. A *GREB1*-specific sgRNA was generated that recognized the sequence GCTGACGGGACGACACATCTG on the sense strand, using a PAM site that is adjacent to the *GREB1* termination codon. For targeting *ESR1*, the reporter gene cassette was subcloned into a construct containing approximately 2 kb of *ESR1* gene surrounding the termination codon in exon 8, such that the reporter gene cassette is located 21 bp downstream of the *ESR1* termination codon in the resulting hybrid mRNA transcript. An *ESR1*-specific sgRNA was generated that recognized the sequence GTCTCCAGCAGCAGGTCATAG on the anti-sense strand, using a PAM site that is 160 bp upstream of the *ESR1* termination codon. Corresponding *Cas9* sgRNA and targeting construct pairs were transiently transfected into MCF7 cells using Lipofectamine 2000 (Thermo Fisher Scientific). After growth for seven days, the cells were cultured in medium containing 5 μg ml^{-1} blasticidin to select for the desired recombinants. Single-cell clones were then isolated by limiting-dilution single-cell cloning in 96-well plates.

Growth rate analysis. Cells were seeded in triplicates in white, clear-bottom, 96-well plates (Corning, 3903), at a density of 5,000 cells per well. Plates were incubated in an IncuCyte ZOOM instrument (Essen Bioscience) at 37 °C, 5% CO_2. Four non-overlapping phase-contrast images (10×) were taken every 2 h for a total of 160 h. IncuCyte ZOOM software (version 2015A) was used to calculate the mean confluence per well at each time point (filtered to exclude objects smaller than 100 μm^2), and averaged across wells to calculate the mean confluence per strain. Doubling times were calculated for each strain, using the formula $T_{doubling} = (\log_2(\Delta T))/(\log(c_2) - \log(c_1))$, in which c_1 and c_2 were the minimum and maximum percentage confluency during the linear growth phase, respectively, and ΔT was the time elapsed between c_1 and c_2. To account for potential differences in cell recovery following seeding, $t = 0$ h was defined as the first time point in which the mean strain confluency surpassed a threshold of 15%. To examine the effect of oestrogen depletion on the growth of MCF7 strains, cells were cultured either in standard conditions (described above) or in oestrogen-depleted conditions: RPMI 1640 without phenol red (Life Technologies), with 10% charcoal-stripped fetal bovine serum (Life Sciences) and 1% penicillin–streptomycin–glutamine (Life Technologies). Comparison between standard and oestrogen-depleted conditions was performed by calculating the fold change in doubling time between the two conditions.

Cell painting. Cells were plated in triplicate at a density of 1,000 cells per well, and then stained and fixed as previously described[26,30]. Images were taken on a Perkin-Elmer Opera Phenix microscope with a 20×/1.0NA water-immersion lens. Image quality control was carried out as previously described[31], using CellProfiler[30] and CellProfiler-Analyst[31]. For all 27 MCF7 strains, the majority of images in all three wells passed quality control, and therefore all strains were further considered. Image illumination correction and analysis were performed in CellProfiler. For each of the 27 MCF7 strains, the median value of the 1,784 measured features was computed and used for hierarchical clustering.

DNA and RNA extraction. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen), according to the manufacturer's protocol. Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen), according to the manufacturer's protocol.

DNA fingerprinting. Fingerprinting analysis was performed using 44 polymorphic loci. 'GenotypeConcordance' (Picard tools) was used to calculate the concordance between each pair of samples (separately for the MCF7 and A549 cohorts). Samples with >95% concordance were called a match.

Ultra-low-pass whole-genome DNA sequencing. Copy number characterization was performed using low-pass (approximately 0.2× coverage) whole-genome sequencing. Libraries were prepared from 50 ng of DNA using ThruPLEX-DNAseq sample preparation kits (Rubicon Genomics) according to the manufacturer's protocol. The resultant libraries were quantified using a Qubit fluorometer (Agilent TapeStation 2200) and RT–qPCR using the Kapa Library Quantification kit (Kapa Biosystems), according to the manufacturer's protocol. Uniquely indexed libraries were pooled in equimolar ratios and sequenced on a single Illumina NextSeq500 run with paired-end 35-bp reads, at the Dana-Farber Cancer Institute Molecular Biology Core Facilities. The reads were aligned to the UCSC hg19 reference genome, using 'BWA-MEM' (v.0.07.15), with default parameters.

Ultra-low-pass whole-genome DNA-sequencing data analysis. The ichorCNA algorithm[32] was applied to identify CNAs of large genomic segments, chromosome arms and whole chromosomes. First, the genome was divided into 1-Mb bins and read counts were generated for each bin using the HMMcopy Suite (http://compbio.bccrc.ca/software/hmmcopy/). The raw read counts were then normalized to correct for GC content and mapability biases using the HMMcopy R package[33], generating corrected read counts for each 1-Mb bin. Segmentation and copy number prediction for each sample were performed using ichorCNA v.0.1.0 (https://github.com/broadinstitute/ichorCNA), which is optimized for low-coverage whole-genome sequencing. Parameters were initialized based on prior knowledge: –normal = 0.01, –ploidy = c(3, 3.5, 4), –txnE = 0.99999, –txnStrength = 10,000, –maxCN = 8. Remaining parameters were set to the default. For bin-level comparison between strains, we used the $\log_2$-transformed corrected read counts and determined gain and loss status using thresholds of 0.1 and −0.1, respectively. For arm-level calls, the copy number status was determined based on the largest overlapping segment.

Deep targeted sequencing. Deep (approximately 250× coverage) targeted exon sequencing of 447 genes that are commonly mutated in cancer was performed (Profile OncoPanel v.3). Prior to library preparation, DNA was fragmented (Covaris sonication) to 250 bp and further purified using Agentcourt AMPure XP beads. Size-selected DNA was ligated to sequencing adaptors with sample-specific barcodes during automated library preparation (SPRIworks, Beckman-Coulter). Libraries were pooled and sequenced on an Illumina Miseq to estimate library concentration based on the number of index reads per sample. Library construction was considered to be successful if the yield was ≥250 ng, and all samples had sufficiently high yields. Normalized libraries were pooled in batches, and hybrid capture was performed using the Agilent Sureselect Hybrid Capture kit with the POPv3_824272 bait set[34]. The list of 447 genes included in POPv3_824272 is provided as Supplementary Table 2. Captures were then pooled and sequenced on one HiSeq3000 lane. Pooled sample reads were deconvoluted and sorted using Picard tools (http://broadinstitute.github.io/picard). The reads were aligned to the reference sequence b37 edition from the Human Genome Reference Consortium using 'bwa aln' (http://bio-bwa.sourceforge.net/bwa.shtml), with the following parameters: '-q 5 -l 32 -k 2 -o 1', and duplicate reads were identified and removed using Picard tools[35]. The alignments were further refined using the GATK tool for localized realignment around indel sites (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_indels_IndelRealigner.php). Recalibration of the quality scores was also performed using GATK tools (http://gatkforums.broadinstitute.org/discussion/44/base-quality-score-recalibration-bqsr)[36,37]. Metrics for the representation of each sample in the pool were generated on the unaligned reads after sorting on the barcode (http://broadinstitute.github.io/picard/picard-metric-definitions.html). All samples achieved the CCGD recommended threshold of >30× coverage for >80% of the targeted bases. Average mean exon target coverage was 251.5× (range: 171.5–336.7×) for the MCF7 samples, 288.9× (range: 208.2–398.9×) for the A549 samples and 257.32 (range 211.7–442.68×) for the additional cell line samples.

Targeted sequencing data analysis. Mutation analysis for SNVs was performed using MuTect v.1.1.4[38]. Indel calling was performed using the SomaticIndelDetector

tool in GATK (http://www.broadinstitute.org/cancer/cga/indelocator). Consecutive variants in the same codon were reannotated to maximize the effect on the codon and marked as 'Phased' variants. MuTect was run in paired mode, pairing the MCF7 or A549 samples to a normal sample, CEPH1408. Mutations were called if detected in >2% of the reads (AF > 0.02). All SNVs, indels and phased variants were annotated with Variant Effect Predictor[39]. Variants were filtered against the 6,500 exome release of the Exome Sequencing Project database. Variants that were represented more than once in either the African- or European–American populations and were found less than twice in COSMIC were considered to be germline (given that no matched normal samples were available). A germline filter was not applied to the downstream analyses, as changes in such mutations between strains of the same cell line would have to arise in culture and may be functionally relevant. Non-silent mutations were considered to be those with the following BestEffect Variant Classification: missense, initiator codon, nonsense, splice acceptor, splice donor, splice region, frameshift, inframe insertion or inframe deletion. Mutations that appeared more than once in COSMIC were regarded as COSMIC mutations. The complete lists of variants (SNVs, indels and phased) for MCF7, A549 and additional cell lines are provided in Supplementary Tables 5, 17 and 23, respectively.

CNAs were identified using RobustCNV, an algorithm that relies on localized changes in the mapping depth of sequenced reads in order to identify changes in copy number at the sampled loci (M. Ducar et al. manuscript in preparation). Systematic bias in mapping depth was reduced using robust regression, fitting the observed tumour mapping depth against a panel of normal samples captured using the same bait set. Observed values were then normalized against predicted values and expressed as $\log_2$ ratios. A second normalization step was performed to remove GC bias, using a LOESS fit. $\log_2$ ratios were centred on segments that were determined to be diploid based on the allele fraction of heterozygous SNPs in the targeted panel. Normalized coverage data were next segmented using Circular Binary Segmentation[40] with the 'DNAcopy' Bioconductor package. Finally, segments were assigned gain, loss or normal-copy calls using a cutoff derived from the within-segment standard deviation of post-normalized mapping depths. Owing to the high data quality and low within-segment standard deviation, a cutoff of around 0.1 was applied to all samples. Segment calls were summarized to gene calls by assigning them to capture intervals, and then counting the interval calls for each gene. Gene level calls were determined according to the following rules: 'gain' = '+' calls >50%; 'loss' = '−' calls >2 or in 100%; 'gain + loss' = '−' calls >2 times and '+' calls <50%; 'mixed' = '+' and '−' calls in the same gene, but below threshold; 'normal+' = '+' calls, but below threshold; "'normal−' = '−' calls, but below threshold; 'Normal' = no '+' or '−' calls. The complete list of CNAs for MCF7, A549 and additional cell lines are provided as Supplementary Tables 4, 16 and 22, respectively.

For a subset of 60 genes (listed in Supplementary Table 2), rearrangements (structural variants) were detected using BreaKmer[41], which is designed to detect larger genomic structural variations from single-sample-aligned short-read target-captured high-throughput sequencing data. In brief, the method extracts 'misaligned' sequences from a targeted region, such as split reads and unmapped mates, assembles a contig from these reads, and re-aligns the contig to make a variant call. It classifies detected variants as 'insertions/deletions', 'tandem duplications', 'inversions' and 'translocations'. The complete lists of structural variants for MCF7 and A549 are provided in Supplementary Tables 6 and 18, respectively. Rearrangements were visualized using the 'Circos' visualization tool[40].

Clonality analysis. To resolve clonal dynamics and composition, we applied the PyClone algorithm v.0.13.0 (https://bitbucket.org/aroth85/pyclone/wiki/Home) to the measured allelic fractions, accounting for copy number, loss of heterozygosity and cellularity[17]. PyClone enabled us to follow clonal dynamics throughout the evolution of cell populations[17,18]. For copy number input, we used results from ichorCNA segmentation and copy number predictions. Mutations with <50 read depth were excluded. The following parameters were used for PyClone: 10,000 iterations, 1,000 burn-in, 'total_copy_number' for the prior. We also performed multi-sample analysis using PyClone, to determine the changes in clonal composition across strains. For the multi-sample analysis, mutations were selected as the union set across all 27 strains. The same parameters were used for PyClone multi-sample analysis as for the individual-sample runs.

DNA barcoding experiment. Degenerate oligonucleotides for sgRNA–barcode library construction were synthesized by IDT and cloned into lentiGuide-Puro[42] by Gibson assembly, as previously described[43]. Approximately 300 μg of Gibson product was transformed into 25 μl of Endura electrocompetent cells (Lucigen). After a 1-h recovery period, 0.1% of transformed bacteria were plated in a tenfold dilution series on ampicillin-containing plates to determine the number of successful transformants. The remainder of the transformed bacteria were cultured in 50 ml of LB with 50 μg ml^{-1} ampicillin for 16 h at 30 °C. Plasmid libraries were extracted using a Plasmid MidiPlus kit (Qiagen) and sequenced to a depth of 6.2 million reads on a Illumina Miniseq, corresponding to 6× coverage of >1 million barcodes. Lentivirus was prepared by transfecting a total of 10 million HEK293FT cells, as previously described[43]. The MCF7-D strain was cultured in standard conditions (described above), and four million cells were infected with a low multiplicity of infection (20–30%) to reduce the probability of each cell being infected with more than one barcode. Cells underwent puromycin selection, and the final cell pool contained approximately 160,000 unique barcodes. Cells were expanded for the experiment, and five million cells were then plated into 25 tissue culture flasks. Five culture conditions were then applied (with five replicates per condition): (1) RPMI 1640 (Life Technologies) with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin–streptomycin–glutamine (Life Technologies); (2) DMEM (Life Technologies) with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin–streptomycin–glutamine (Life Technologies); (3) RPMI 1640 without phenol red (Life Technologies), with 10% charcoal-stripped fetal bovine serum (Life Sciences) and 1% penicillin–streptomycin–glutamine (Life Technologies); (4) RPMI 1640 (Life Technologies) with 10% fetal bovine serum (Sigma-Aldrich), 1% penicillin–streptomycin–glutamine (Life Technologies) and 0.05% DMSO (Sigma-Aldrich); (5) RPMI 1640 (Life Technologies) with 10% fetal bovine serum (Sigma-Aldrich) and 1% penicillin–streptomycin–glutamine (Life Technologies), supplemented for the first 48 h with 500 nM bortezomib (Selleckchem, S1013). After five weeks of culture, DNA was extracted and barcode abundance was assessed by DNA sequencing, as previously described[43]. Libraries were sequenced to a median depth of 4.2 million reads, corresponding to a barcode coverage of >26×.

Transcriptional profiling with L1000. The L1000 expression-profiling assay was performed as previously described[16]. First, mRNA was captured from cell lysate using an oligo dT-coated 384-well Turbocapture plate. The lysate was then removed, and a reverse-transcription mix containing MMLV was added. The plate was washed and a mixture containing both upstream and downstream probes for each gene was added. Each probe contained a gene-specific sequence, along with a universal primer site. The upstream probe also contained a microbead-specific barcode sequence. The probes were annealed to the cDNA over a 6-h period, and then ligated together to form a PCR template. After ligation, Hot Start *Taq* and universal primers were added to the plate. The upstream primer was biotinylated to allow later staining with streptavidin–phycoerythrin. The PCR amplicon was then hybridized to Luminex microbeads via the complimentary, probe-specific barcode on each bead. After overnight hybridization the beads were washed and stained with streptavidin–phycoerythrin to prepare them for detection in Luminex FlexMap 3D scanners. The scanners measured each bead independently and reported the bead colour and identity and the fluorescence intensity of the stain. A deconvolution algorithm converted these raw fluorescence intensity measurements into median fluorescence intensities for each of the 978 measured genes, producing the GEX level data. These GEX data were then normalized based on an invariant gene set, and then quantile-normalized to produce QNORM level data. An inference model was applied to the QNORM data to infer gene expression changes for a total of 10,174 features. Per-strain gene expression signatures were calculated using a weighted average of the replicates, for which the weights are proportional to the Spearman correlation between the replicates.

Transcriptional profiling data analysis. To examine how newly profiled MCF7 and A549 cells compared in gene expression to a previously acquired collection of cell line profiles (untreated samples that served as controls for connectivity map perturbational experiments), we used *t*-SNE. Profiles were restricted to untreated profiles from the nine core connectivity map cell lines, and to batches with multiple untreated profiles. Because samples were first clustered based on their project codes, batch effect was next removed using the COMBAT algorithm[44]. *t*-SNE analysis was applied to the batch-corrected data and visualized using a scatter plot. Analysis was completed using the 'Rtsne' R package version 0.13[62]. For the comparison of transcriptional variation across the nine core connectivity map cell lines, the collection of untreated profiles generated with the L1000 assay was used. Five profiles from each cell line were randomly chosen, and the expression variance of the 978 L1000 landmark genes was calculated for each cell line. For the comparison of L1000 gene expression data to the Cancer Cell Line Encyclopedia (CCLE) gene expression profiles, RNA-sequencing (RNA-seq) and Affymetrix gene expression profiles were downloaded from the CCLE website (https://portals.broadinstitute.org/ccle/data). Data within each platform were processed using invariant set scaling, which adjusts profiles according the expression of 80 'invariant' genes, followed by quantile normalization[16]. The ranked gene expression order of the 978 landmark genes was compared using a Spearman's correlation.

Chemical screening. MCF7 strains were tested against a small-molecule informer set library of 321 anti-cancer compounds, assembled by the Cancer Target Discovery and Development (CTD^2; https://ocg.cancer.gov/programs/ctd2/data-portal), using the same principles as those described in the Cancer Therapeutics Response Portal[8,45]. The list of screened compounds is included as Supplementary Table 26. Cells were seeded in their culture medium in white, 384-well plates (Corning, 3570) at an initial density of 2,500 cells per well and incubated overnight at 37 °C, 5% CO_2. The next day, 25 nl (for primary screen) or 100 nl

(for confirmation dose–response screen) of compound stocks in DMSO were added by pin transfer. Plates were incubated for 72 h, cooled at room temperature for 10 min, and viability was measured using the CellTiter-Glo luminescent cell viability assay (Promega), according to the manufacturer's protocol. After 10 min of incubation, luminescence was read on a Perkin Elmer Envision reader, at a speed of 0.1 s per well.

Chemical screening data analysis. Data were analysed in Genedata Screener version 13.0, using the normalization method 'neutral controls', for which the median of the 32 DMSO wells on each plate was set to 0% activity and 0 raw signal was set to −100%. Positive controls (20 μM MG-132 or 20 μM bortezomib) were included on all plates (16 wells each) but were not used for normalization owing to variability in the response across cell lines. Dose–response curves were fit using the 'Smart Fit' strategy in Genedata. The percentage of effect was defined as the high-concentration asymptote (sinf) and the qEC_{50} was the concentration at which the fitted curve crossed the inhibitory value representing half of the maximal effect (%). In addition, parameters were calculated at which the curve crossed absolute inhibitory values of 30% or 50% regardless of maximal effect ($AbsEC_{30}$ and $AbsEC_{50}$, respectively). AUC calculations were performed as previously described[8]: curves were fit with nonlinear sigmoid functions, forcing the low concentration asymptote to 1 using a three-parameter sigmoidal curve fit. The AUC for each compound–strain pair was calculated by numerically integrating under the eight-point concentration–response curve. For visualization purposes, drug response curves were fit with a four-parameter log-logistic function, based on normalized viability data from which the lowest dose viability had been subtracted. Plots were generated using the "LL.4" function in the 'drc' R package (https://cran.r-project.org/web/packages/drc/). To examine a potential link between proliferation rate and differential drug response, doubling times were compared against the AUC values of the 33 differentially-active compounds.

Gene set enrichment analysis. Gene set enrichment analysis (GSEA) was performed using the 10,147 genes best inferred from the connectivity map linear model[33], also known as the BING gene set. Samples were divided into two classes depending on the comparisons being made: samples with a genetic alteration versus samples without it; samples sensitive to a drug (>50% inhibition) versus samples insensitive to the same drug (<20% inhibition). Differential expression was calculated using the signal-to-noise metric[46]. A ranked gene list and signal-to-noise values served as the input for the GSEA preranked module of GSEA, using the Java app version 3.0. The analysis was run using the 'hallmark', 'KEGG', 'positional' and 'oncogenic' signature collections from the Molecular Signature Database (MsigDB)[27] (http://software.broadinstitute.org/gsea/msigdb). To compare between our MCF7 panel, CTD[2] and GDSC, drug responses were downloaded from the CTRP website (https://ocg.cancer.gov/programs/ctd2/data-portal; 'v20.data.curves_post_qc' file, updated 14 October 2015) and from the GDSC website (http://www.cancerrxgene.org/downloads; 'log(IC_{50}) and AUC values' file, updated 4 July 2016). Expression profiles were downloaded from the CCLE website to match the CTD[2] drug-response data (https://portals.broadinstitute.org/ccle/data; 'CCLE_Expression_Entrez_2012-09-29.gct', updated 17 October 2012), and from the GDSC website to match with the GDSC drug response data (http://www.cancerrxgene.org/downloads; 'RMA normalized expression data for cell lines', updated 2 March 2017). Expression profiles were filtered to include only the genes that belong to the L1000 BING set. GSEA compared the expression patterns of the five strains or cell lines with the highest AUC values for each matched drug with the five strains/cell lines with the lowest AUC values for that drug. As the robustness of gene expression signatures varies, this quantitative analysis was restricted to the 50 well-defined hallmark GSEA gene sets[27].

Single-cell RNA-seq. MCF7 cells were cultured as described above. To follow transcriptional changes after drug treatment, MCF7-AA cells were exposed to 500 nM of bortezomib (Selleckchem, S1013) and collected before treatment, after 12 h of exposure (t_{12}), after 48 h of exposure (t_{48}) or after 72 h of exposure followed by drug wash and 24 h of recovery (t_{96}). Cells were washed, trypsinized, passed through a 40-μm cell strainer, centrifuged at 400*g* and resuspended at a concentration of 1,000 cells per μl in PBS containing 0.5% BSA. Single cells were processed through the Chromium Single Cell 3′ Solution platform using the Chromium Single Cell 3′ Gel Bead, Chip and Library Kits (10X Genomics) per the manufacturer's protocol. In brief, 7,000 cells were added to each channel and were then partitioned into Gel Beads in emulsion in the Chromium instrument, where cell lysis and barcoded reverse transcription of RNA occurred, followed by amplification, shearing and 5′ adaptor and sample index attachment. Libraries were sequenced on an Illumina NextSeq 500.

Single-cell RNA-seq data analysis. Reads were mapped to the GRCh38 human transcriptome using cell ranger version 2.1.0, and transcript-per-million values were calculated for each gene in each filtered cell barcoded sample. Transcript-per-million values were then divided by 10, since the complexity of single-cell libraries is estimated to be in the order of 100,000 transcripts. For each cell, we quantified the number of expressed genes and the proportion of the transcript counts derived from mitochondrial genes. Cells with either <1,000 detected genes or >0.15 mitochondrial fraction were excluded from further analysis. Finally, the resulting expression matrix was filtered to remove genes detected in <3 cells. We focused on highly variable genes for downstream principal component analysis (PCA). For each dataset, we used the Seurat[47] (http://satijalab.org/seurat/) R package to detect variable genes based on fitting a relationship between the mean and the dispersion of each gene. We next scaled the data and regressed out unique molecular identification number and mitochondrial gene fraction to remove technical noise. The resulting scaled data were used as an input for PCA. Top significant principal components, estimated by a manual inspection of the PCA standard deviations elbow plots, were used to generate *t*-SNE plots. For each dataset, we used Seurat[47] (http://satijalab.org/seurat/) to identify genes that vary between samples. To detect differentially active pathways, gene ontology (GO) enrichment analysis was performed with MSigDB[27] (http://software.broadinstitute.org/gsea/msigdb) using the differentially expressed genes that passed the following thresholds: $|\log_2(\text{fold change})| > 0.25$, Bonferroni-corrected $P < 0.01$, the gene was detected in >10% of the cells in each of the compared groups. Expression signatures for selected pathways were downloaded from MSigDB[27]. We evaluated the degree to which individual cells express a certain expression signature by using a procedure that takes into account the variability in signal-to-noise ratio, as previously reported[48]. To calculate pairwise cell distances, variable genes were detected, and the cell embedding matrix for the top significant principal components was used to calculate the Euclidean distance between every two cells within each sample.

Analysis of genome-wide CRISPR screens. CERES dependency scores[49] were obtained from the Broad Institute Achilles website (https://portals.broadinstitute.org/achilles/datasets/18/download). Owing to an unusually large difference in screen quality between MCF7 and KPL1, the subtle differences in dependency status between these lines were dominated by effects related to screen quality. To remove these uninteresting sources of variation, we corrected CERES gene scores by removing their first six principal components. These components were well-explained by experimental batch effects related to screen performance and plasmid DNA pool. Corrected dependency scores <−0.5 were defined as dependencies. Genes listed as 'pan_dependent' in the original dependency dataset were excluded from further analysis. For a more stringent overlap comparison, genes with CERES scores between −0.4 and −0.6 in MCF7 or KPL1 were further excluded. To implement the force-directed layout, described in Extended Data Fig. 11b, the full corrected dependency matrix was reduced to its top 100 principle components and a *k*-means clustering algorithm was run repeatedly on cell lines. Here, *k* is the number of clusters, and the mean cluster size (number of cell lines) divided by *k* is a parameter similar to perplexity in *t*-SNE, set to 6 for our data. Edges between cells were weighted according to the frequency with which they clustered together, with edges appearing less than 30% of the time ignored. Cells were then laid out using the SFPD spring-block algorithm[50]. Cell line RNA-seq gene expression data and reverse-phase protein array protein expression data were obtained from the CCLE website (https://portals.broadinstitute.org/ccle/data). Single-sample GSEA was calculated using the ssGSEA algorithm[51].

Chymotrypsin-like activity. MCF7 cells were plated in triplicates in 96-well plates at a density of 20,000 cells per well. After 24 h, chymotrypsin-like activity of the proteasome was assayed, using the Proteasome-Glo assay (Promega), according to manufacturer's protocol. The activity levels were normalized to the relative cell number that was measured using the fluorescent detection of resazurin dye reduction (544-nm excitation and 590-nm emission).

Western blots. For PSMC2 and PSMD2 immunoblotting, cells were lysed in HENG buffer (50 mM HEPES-KOH pH 7.9, 150 mM NaCl, 2 mM EDTA pH 8.0, 20 mM sodium molybdate, 0.5% Triton X-100, 5% glycerol), with protease inhibitor cocktail (Roche Diagnostics, 11836153001). Protein concentrations were determined by BCA assay (Thermo Fisher Scientific, 23227) and proteins were resolved using SDS–PAGE for immunoblot analysis. Antibodies against the following human proteins were used: α-tubulin (ab80779, Abcam), PSMC2 (MSS1-104, Enzo Life Sciences) and PSMD1 (C-7, Santa-Cruz). Visualization was performed using the ChemiDoc MP System (Bio-Rad) and ImageLab Software (Bio-Rad) was used to quantify relative band intensities. For ERα immunonblotting, cells were lysed with a mix of 4× protein loading buffer (Li-Cor, 928-40004) and 10× NuPAGE sample reducing agent (Life Technologies, NP0009). Protein concentration was normalized by cell counting and proteins were resolved by SDS–PAGE. Antibodies against the following human proteins were used: β-actin (N-21, Santa Cruz), ERα (F-10, Santa Cruz). Visualization was performed using the Odyssey CLx imaging machine (Li-Cor) and Image Studio Software (Li-Cor) was used to quantify the relative intensities.

Generation and comparison of dendrograms. Dendrograms were constructed using Euclidean distances for continuous measures and Manhattan distances for discrete measures. Complete linkage hierarchical clustering was performed in all cases. The mutation status dendrogram was based on mutations with AF > 0.05. The gene expression dendrogram was based on the 978 landmark genes directly

measured by the L1000 assay. The copy number dendrograms were based on discrete calls (loss, normal or gain) assigned to each event based on its $\log_2$ copy number ratio, using a cutoff value of ± 0.1. The drug-response dendrogram was based on normalized viability values. The cell morphology dendrogram was based on the full list of the 1,784 measured cellular features. The barcode representation dendrogram was based on the $\log_2$ transformed number of reads, including only barcodes with >1,000 reads in at least one sample. To understand how dendrograms from different sources compared, the Fowlkes–Mallows index was used, as it could capture similarities in global clustering while ignoring within-group variance[52]. The 'Bk' function in the 'dendextend' R package was used for computations and visualizations. We compared dendrograms from different sources with k values ranging from 5 to 26. A background distribution was calculated by randomly shuffling the labels of the trees 1,000 times, and calculating Bk values. The 95% upper quantile of the randomized distribution for each k was plotted. The maximum Bk value was used to estimate the degree of similarity between the compared pair of dendrograms.

Calculation of the distances between strains based on their genomic features. CNA distance based on ultra-low-pass whole-genome DNA sequencing was determined by the fraction of the genome affected by discordant CNA calls. CNA and SNV distances based on targeted sequencing were determined by Jaccard indices, defined as the number of shared events between strains (intersection) divided by the total number of events in these strains (union). For SNVs, both the mutated gene and the exact amino acid change had to be identical to be counted as a shared event. Gene expression distances were defined as the Euclidean distances between L1000 expression profiles. Drug-response distances were defined as the Euclidean distances between drug-response profiles, after limiting the drug set to active drugs only (that is, drugs that reduced the viability of at least one strain by >50%) and setting the threshold for viability values to ± 100.

Comparisons across CCLE cell lines. Gene-level mRNA expression, copy number and mutation status data were downloaded from the CCLE website (https://portals.broadinstitute.org/ccle/data; 'CCLE_Expression_Entrez_2012-09-29.gct', updated 17 October 2012; 'CCLE_copynumber_byGene_2013-12-03.txt', updated 27 May 2014; 'CCLE_MUT_CNA_AMP_DEL_binary_Revealer.gct', updated 29 February 2016). The total number of point mutations and copy number changes were counted for each cell line. Chromosome arm-level events in CCLE samples were generated as previously described[53], and the number of arm-level events was counted for each cell line. The fraction of the genome affected by subclonal events was estimated using ABSOLUTE[54]. Combined CNA–SNV genomic instability scores were calculated as described previousy[55]. The DNA repair gene set was derived from MSigDB (http://software.broadinstitute.org/gsea/msigdb), using the 'DNA_Repair' GO signature[56]. The CIN70 gene set was derived from a previous publication[57]. For each gene set, genes that were not expressed at all in the CCLE dataset were removed, and the remaining gene expression values were $\log_2$-transformed and scaled by subtracting the gene expression means. The signature score was defined as the sum of these scaled gene expression values.

Comparison of Broad (CCLE) and Sanger (GDSC) genomic features. Whole-exome sequencing data for 107 matched cell lines were downloaded from the Sanger Institute (http://cancer.sanger.ac.uk/cell_lines, EGA accession number: EGAD00001001039) for the GDSC cell lines, and from the GDC portal (https://portal.gdc.cancer.gov/legacy-archive) for the CCLE cell lines. For copy number analysis, the GATK4 somatic copy number variant pipeline was applied (https://gatkforums.broadinstitute.org/gatk/discussion/9143/how-to-call-somatic-copy-number-variants-using-gatk4-cnv)[36,37]. Gene-level copy number calls were generated by mapping genes from segment calls using the Consensus Coding Sequence database[58]. The gene-level values were $\log_2$ transformed, and converted to discrete values using predefined thresholds (± 0.1, ± 0.3 and ± 0.5). To determine the percentage of discordance for each cell line, the number of discordant CNA calls between each pair of strains was divided by the total number of genes (excluding genes with a neutral copy number call in both datasets). For analysis of somatic variants, the CCLE–Sanger merged mutation calls were downloaded from the CCLE portal (https://portals.broadinstitute.org/ccle/data), and target interval list files were generated for each of the 107 matched cell lines in CCLE. Mutation calling was performed using MuTect[38], with default parameters and '–force_output' enabled, to count the number of reads supporting the reference and alternate allele for each variant in each cell line. For analysis of germline variants, a common target interval list file that consisted of a panel of 105,995 SNPs was generated, based on common SNVs found in 1,019 CCLE RNA-seq samples, and Mutect was applied with the same parameters as described above. Comparison of allelic fractions was performed using the subset of variants with minimum depth of coverage of 10 in both Sanger and CCLE datasets and with minimum of allelic fraction of 0.1 in at least one dataset. Out of the 107 cell lines, one cell line (Dov13) lacked any germline concordance and was thus excluded from all analyses.

Cytogenetic analysis. Karyotyping was performed by KaryoLogic (www.karyologic.com/) on 50 G-banded metaphase spreads per sample. Every spread displayed multiple chromosomal rearrangements with many marker chromosomes. A marker was defined as 'a structurally abnormal chromosome that cannot be unambiguously identified by conventional banding cytogenetics'. The analysis was performed according to the International System for Human Cytogenetic Nomenclature (ISCN) 2016 guidelines. Rare metaphases with >100 chromosomes were excluded from further analysis.

e-karyotyping analysis. RNA-seq data from non-manipulated and non-treated samples of the near-diploid human cell line RPE1 were downloaded from the NCBI SRA website (https://www.ncbi.nlm.nih.gov/sra). STAR-paired aligner was used to align paired-end samples, and STAR-non-paired aligner was used to align the non-paired samples[59]. The STAR to RSEM tool[60] was used to generate the gene-level expression values using the GTEx pipeline (https://github.com/broadinstitute/gtex-pipeline). To infer arm-level copy number changes from gene expression profiles, the RSEM values were analysed using the e-karyotyping method[61]. In brief, RSEM values were $\log_2$-converted, genes that were not expressed ($\log_2$RSEM < 1) in >20% of the samples were excluded, and expression levels of the remaining genes were floored to RSEM = 1. The median expression value of each gene across all samples was subtracted from the expression value of that gene, in order to obtain comparative values. The 10% most variable genes were removed from the dataset to reduce transcriptional noise. The relative gene expression data were then subjected to a CGH–PCF analysis, with a stringent set of parameters: Least allowed deviation = 0.5; Least allowed aberration size = 30; Winsorize at quantile = 0.001; Penalty = 18; Threshold = 0.01. CNAs exceeding 80% of the length of a chromosome arm were called arm-level CNAs.

Comparison of arm-level CNAs between cell line propagation and tumour progression. Recurrence of chromosome arm-level CNAs during breast cancer progression was determined by their frequency in TCGA samples, as previously described[53]. Recurrence of chromosome arm-level CNAs during cell line propagation was determined by comparing the arm-level calls of the strains directly separated by extensive passaging (strain D versus strain L versus strain AA, strain B versus strains I and P), as shown in Extended Data Fig. 2a. Only arms that are recurrently gained or lost (but not both) in TCGA ($Q < 0.05$), and that have variable copy number status across the MCF7 panel, were considered for the comparison.

Statistical analysis. The significance of the difference between genomic instability associated with different sources of genetic variation and difference between chromosome numbers at two time points of single-cell-derived clones was determined using the two-tailed Wilcoxon rank-sum test. The significance of the difference in the Euclidean distance between compounds that work through the same MoA and compounds that work through different MoA's, in the discordance of non-silent SNVs at different stages of transformation, in chromosomal instability (CIN70) and weighted-genomic integrity index (wGII) scores between cell lines derived from primary tumours and those derived from metastases, between the somatic and germline SNV Pearson correlations of the Broad-Sanger cell lines, in the Broad-Sanger somatic SNV concordance between microsatellite-stable and microsatellite-unstable cell lines and between primary tumour-derived and metastasis-derived cell lines was determined using a one-tailed Wilcoxon rank-sum test. The significance of the difference in mutation cellular prevalence across strains was determined by Kruskal–Wallis test. The significance of the difference in AKT inhibitor IV sensitivity between $PTEN^{WT}$ and $PTEN^{MUT}$ strains, in the relative growth effect of ER depletion between *ESR1* loss and no-*ESR1* loss strains, in proteasome activity between bortezomib-sensitive and bortezomib-insensitive strains, in ERα protein expression levels between strains and in the number of arm-level CNAs between matched early-late MCF7 strains was determined using a one-tailed Student's t-test. The significance of the difference in doubling times and in sensitivity to oestrogen depletion was determined using a two-tailed Student's t-test. The significance of the correlation between the two replicates of the primary screen was determined using Pearson's correlation. The significance of the correlation between doubling time and the number of protein coding mutations, the correlation between doubling time and the fraction of subclonal mutations, the correlation between doubling time and drug response, were determined using a Spearman's correlation, excluding the broadly resistant strains Q and M. The significance of the correlation between *ESR1* CERES dependency scores and oestrogen signalling and between *GATA3* CERES dependency scores and GATA3 protein expression levels was determined using a Spearman's correlation. The deviation of the doubling-time–drug-response correlations from a hypothetical mean value of 0 was determined using a two-tailed one-sample t-test. The significance of the difference between the emergence and disappearance of recurrent arm-level CNAs during cell line propagation was determined using McNemar's test. The significance of the correlation between the primary and secondary drug screens was determined using a Spearman's correlation (including only compounds that were active in both screens). The significance of the directionality of drug-pathway association, and the likelihood that a mutation would be clonal given the number of reads that detected it, were determined using a binomial test. The significance between

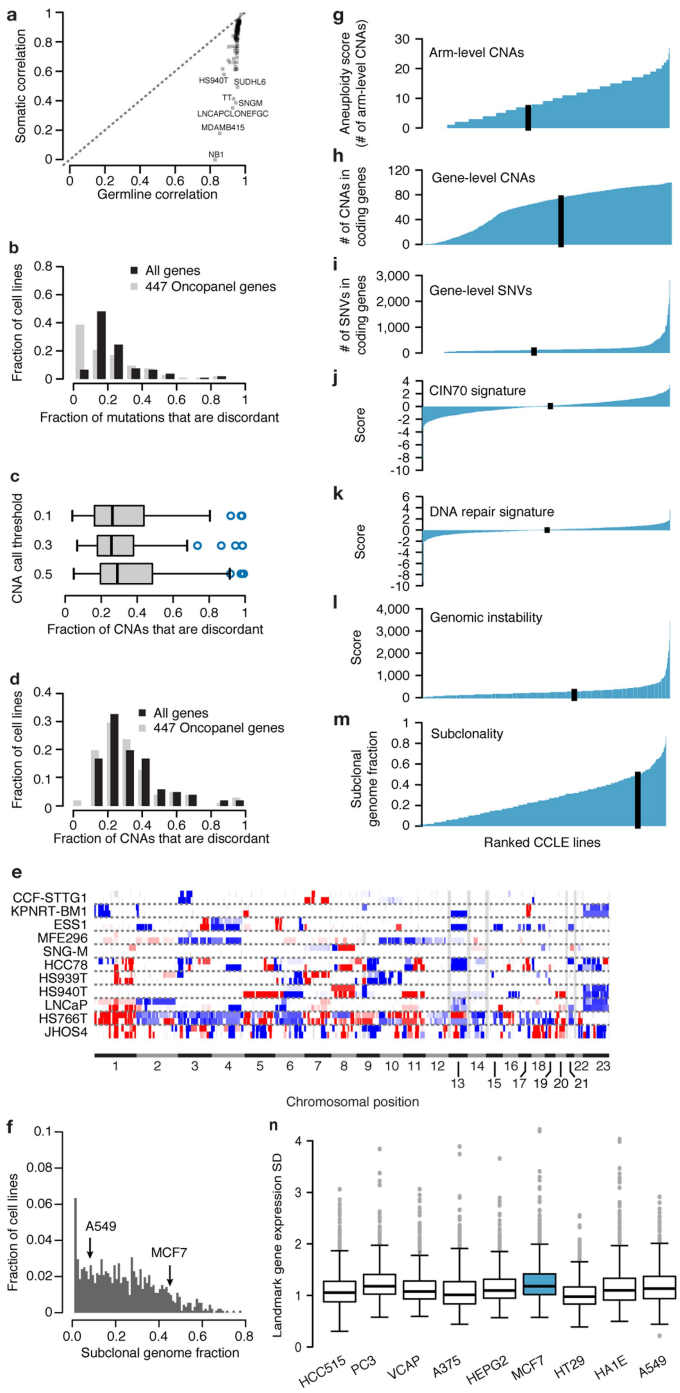
the fraction of pathways correctly identified between the MCF7 panel, CTD^2 and GDSC was determined using a two-tailed Fisher's exact test. GSEA P values and FDR-corrected Q values are shown as provided by the default analysis output. For the comparison of pathway prediction shown in Supplementary Table 32, FDR Q values were recalculated using only the pre-selected pathways. Thresholds for significant associations were determined as: $P < 0.05$; $Q < 0.25$. The significance of the difference in the karyotypic variation between parental and single-cell-clone derived cultures was determined using the Levene's test. The significance of differentially expressed genes in the single-cell RNA-seq data was determined by an analysis of variance (ANOVA) followed by a Games–Howell post hoc test and a Bonferroni correction. Box plots show the median, 25th and 75th percentiles, lower whiskers show data within 25th percentile $-1.5\times$ the interquartile range (IQR), upper whiskers show data within 75th percentile $+1.5\times$ the IQR, and circles show the actual data points. Statistical tests were performed using the R statistical software (http://www.r-project.org/), and the box plots and violin plots were generated using the 'boxplot' and 'vioplot' R packages, respectively.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

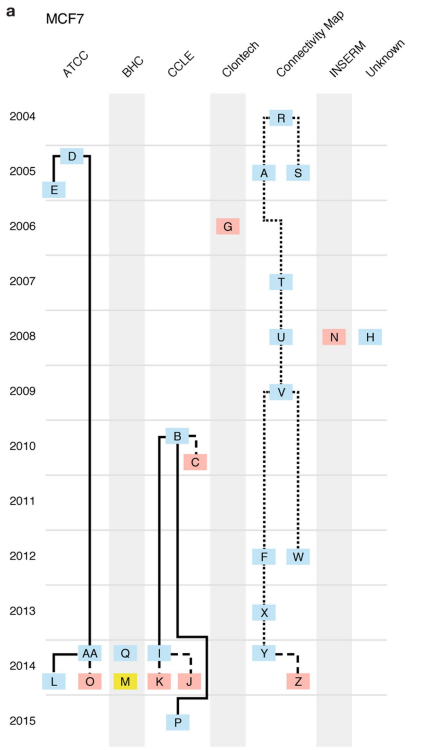
Code availability. The code used to generate and/or analyse the data during the current study are publicly available, or available from the corresponding authors upon request.

Data availability. The datasets generated during and/or analysed during the current study are available within the article, its Supplementary Information or from the corresponding authors upon request. DNA sequencing data were deposited to SRA with BioProject accession number PRJNA398960. Single-cell RNA-seq data were deposited to the Gene Expression Omnibus (GEO, accession number GSE114462). Source Data of all immunostaining blots are available in the online version of this paper. The cell divergence portal is accessible at: https://cellstrainer.broadinstitute.org.

30. Bray, M. A., Fraser, A. N., Hasaka, T. P. & Carpenter, A. E. Workflow and metrics for image quality control in large-scale high-content screens. *J. Biomol. Screen.* **17**, 266–274 (2012).
31. Dao, D. et al. CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* **32**, 3210–3212 (2016).
32. Adalsteinsson, V. A. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).
33. Ha, G. et al. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res.* **22**, 1995–2007 (2012).
34. Sholl, L. M. et al. Institutional implementation of clinical tumor profiling on an unselected cancer population. *JCI Insight* **1**, e87062 (2016).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
37. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
38. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
39. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* **26**, 2069–2070 (2010).
40. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
41. Abo, R. P. et al. BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic Acids Res.* **43**, e19 (2015).
42. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
43. Joung, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
44. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
45. Rees, M. G. et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat. Chem. Biol.* **12**, 109–116 (2016).
46. Golub, T. R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
47. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
48. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
49. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
50. Hu, Y. Efficient, high-quality force-directed graph drawing. *Math. J.* **10**, 37–71 (2006).
51. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
52. Fowlkes, E. B. & Mallows, C. L. A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.* **78**, 553–569 (1983).
53. Ben-David, U. et al. Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat. Genet.* **49**, 1567–1575 (2017).
54. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
55. Zhang, S., Yuan, Y. & Hao, D. A genomic instability score in discriminating nonequivalent outcomes of BRCA1/2 mutations and in predicting outcomes of ovarian cancer treated with platinum-based chemotherapy. *PLoS ONE* **9**, e113169 (2014).
56. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
57. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **38**, 1043–1048 (2006).
58. Pujar, S. et al. Consensus coding sequence (CCDS) database: a standardized set of human and mouse protein-coding regions supported by expert curation. *Nucleic Acids Res.* **46**, D221–D228 (2018).
59. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
60. Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493–500 (2010).
61. Ben-David, U., Mayshar, Y. & Benvenisty, N. Virtual karyotyping of pluripotent stem cells on the basis of their global gene expression profiles. *Nat. Protoc.* **8**, 989–997 (2013).
62. Krijthe, J. H. *Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation* https://github.com/jkrijthe/Rtsne (2015).

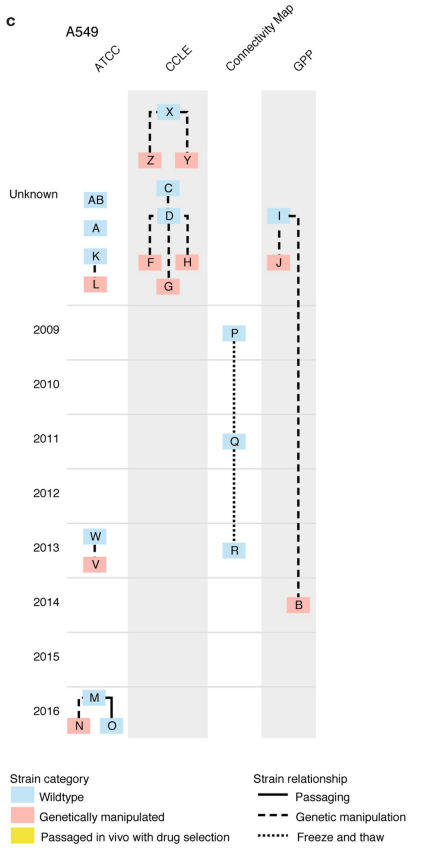


Extended Data Fig. 1 | Comparison of Broad and Sanger genomic features across 106 cell lines. a, Comparison of the Pearson correlations of germline versus somatic SNVs across 106 paired cell lines. **b**, A histogram of the distribution of mutation discordance fractions across cell lines. Black, the distribution of all non-silent SNVs; grey, the distribution of the 447 genes included in the Oncopanel. **c**, Comparison of the fraction of discordant gene-level CNAs between the Broad and the Sanger ($n = 106$ cell lines) datasets, using three different thresholds for CNA calling. Bar, median; box, 25th and 75th percentiles; whiskers, 1.5× IQR of lower and upper quartile; circles, data points. **d**, A histogram of the distribution of CNA discordance fractions across cell lines. Bars are coloured as in **b**. **e**, CNA landscapes of 11 paired cell lines. For each cell line, the CNA landscape of the Broad strain (top) and the Sanger strain (bottom) are shown. Red, copy number gains; blue, copy number losses. CNAs < 10Mb in size are not presented. **f**, A histogram of the fraction of the genome affected by subclonal events across 916 cell lines from the CCLE. MCF7 and A549 are denoted by arrows. **g**, All CCLE cell lines ranked by their aneuploidy scores. **h**, All CCLE cell lines ranked by the number of their gene-level CNAs. **i**, All CCLE cell lines ranked by the number of their gene-level SNVs. **j**, All CCLE cell lines ranked by their chromosomal instability (CIN70) signature scores[57]. **k**, All CCLE cell lines ranked by their DNA-repair signature scores[56]. **l**, All CCLE cell lines ranked by their genomic instability scores[55]. **m**, All CCLE cell lines ranked by their subclonal genome fraction[54]. The vertical black line shows the rank of MCF7 in each comparison. **n**, Comparison of gene expression variation across multiple strains of nine cell lines, including MCF7. Box plots are the standard deviations of the expression levels for the 978 landmark genes directly measured in L1000. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile; circles, all data points.



b

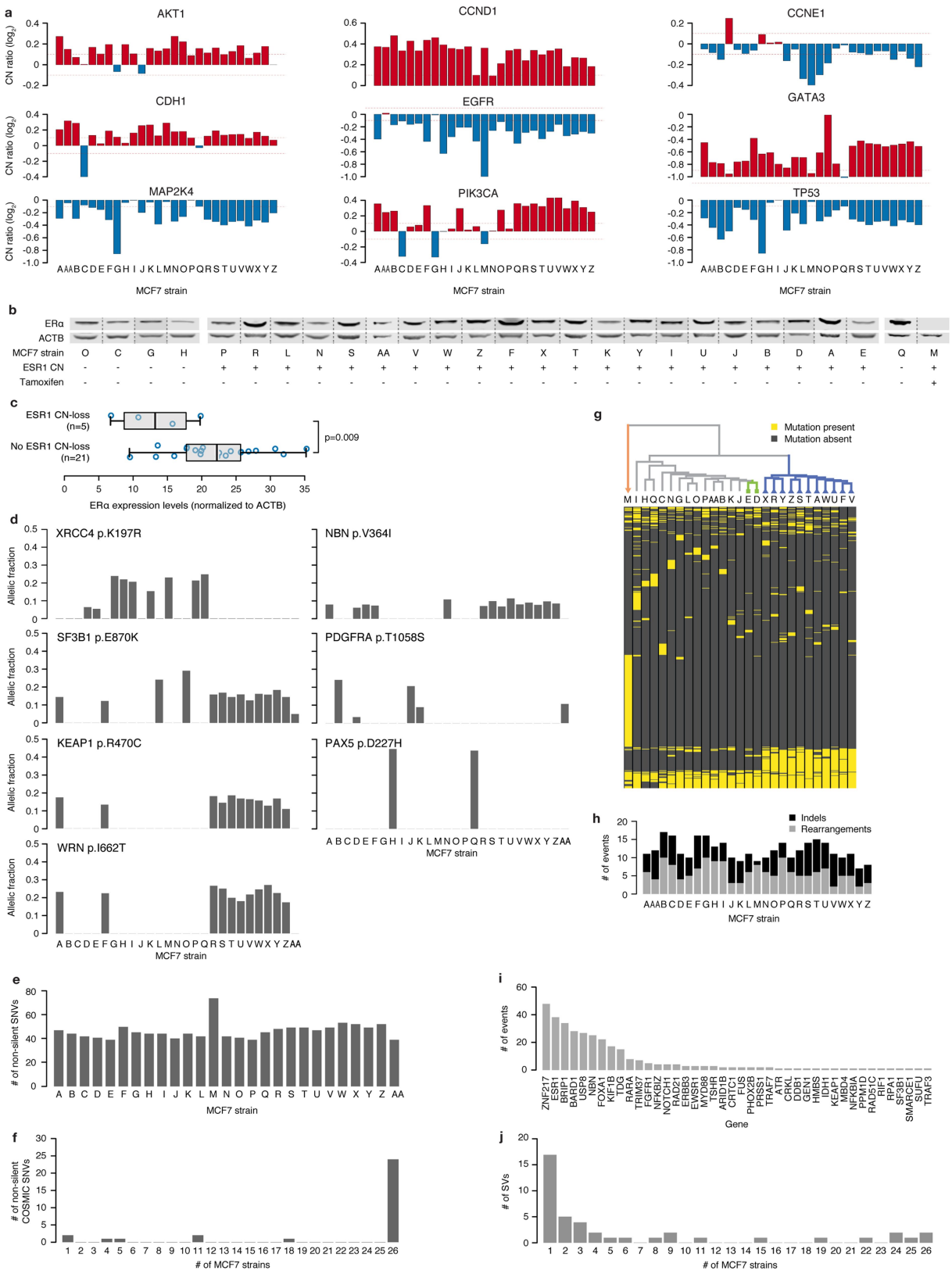
Strain_ID	Origin	Year	Passage	Manipulations	Remarks
MCF7-A	Connectivity Map	2005		Freeze & thaw	
MCF7-B	CCLE	2010			Parental of MCF7-I
MCF7-C	CCLE	2010		EGFP reporter	
MCF7-D	ATCC	2005	p+5		Parental of MCF7-E
MCF7-E	ATCC	2005	p+12		Derived directly from MCF7-D
MCF7-F	Connectivity Map	2012		Freeze & thaw	
MCF7-G	Clontech	2006		Tet-off	
MCF7-H	unknown	2008			
MCF7-I	CCLE	2014		Extensive passaging	Derived directly from MCF7-B; parental of MCF7-K
MCF7-J	CCLE	2014		DNA-barcoded (PRISM)	
MCF7-K	CCLE	2014		Cas9-expressing	Derived directly from MCF7-I
MCF7-L	ATCC	2014	High passage	Extensive passaging	Derived directly from MCF7-AA
MCF7-M	BHC	2014		*In vivo* tamoxifen treatment	Persistent cells: passaged in xenografts --> treated with tamoxifen --> passaged in culture again; derived from same parental as MCF7-Q
MCF7-N	INSERM	2008		YFP reporter	
MCF7-O	ATCC	2014		Luciferase reporter	
MCF7-P	CCLE	2015		Extensive passaging	Derived directly from MCF7-B
MCF7-Q	BHC	2014			Parental of MCF7-M (with continued passaging)
MCF7-R	Connectivity Map	2004			Parental of MCF7-A/F/S/T/U/V/W/X/Y/Z
MCF7-S	Connectivity Map	2005		Freeze & thaw	
MCF7-T	Connectivity Map	2007		Freeze & thaw	
MCF7-U	Connectivity Map	2008		Freeze & thaw	
MCF7-V	Connectivity Map	2009		Freeze & thaw	
MCF7-W	Connectivity Map	2012		Freeze & thaw	
MCF7-X	Connectivity Map	2013		Freeze & thaw	
MCF7-Y	Connectivity Map	2014		Freeze & thaw	
MCF7-Z	Connectivity Map	2014		Cas9-expressing	
MCF7-AA	ATCC	2014	Low passage		Parental of MCF7-L



d

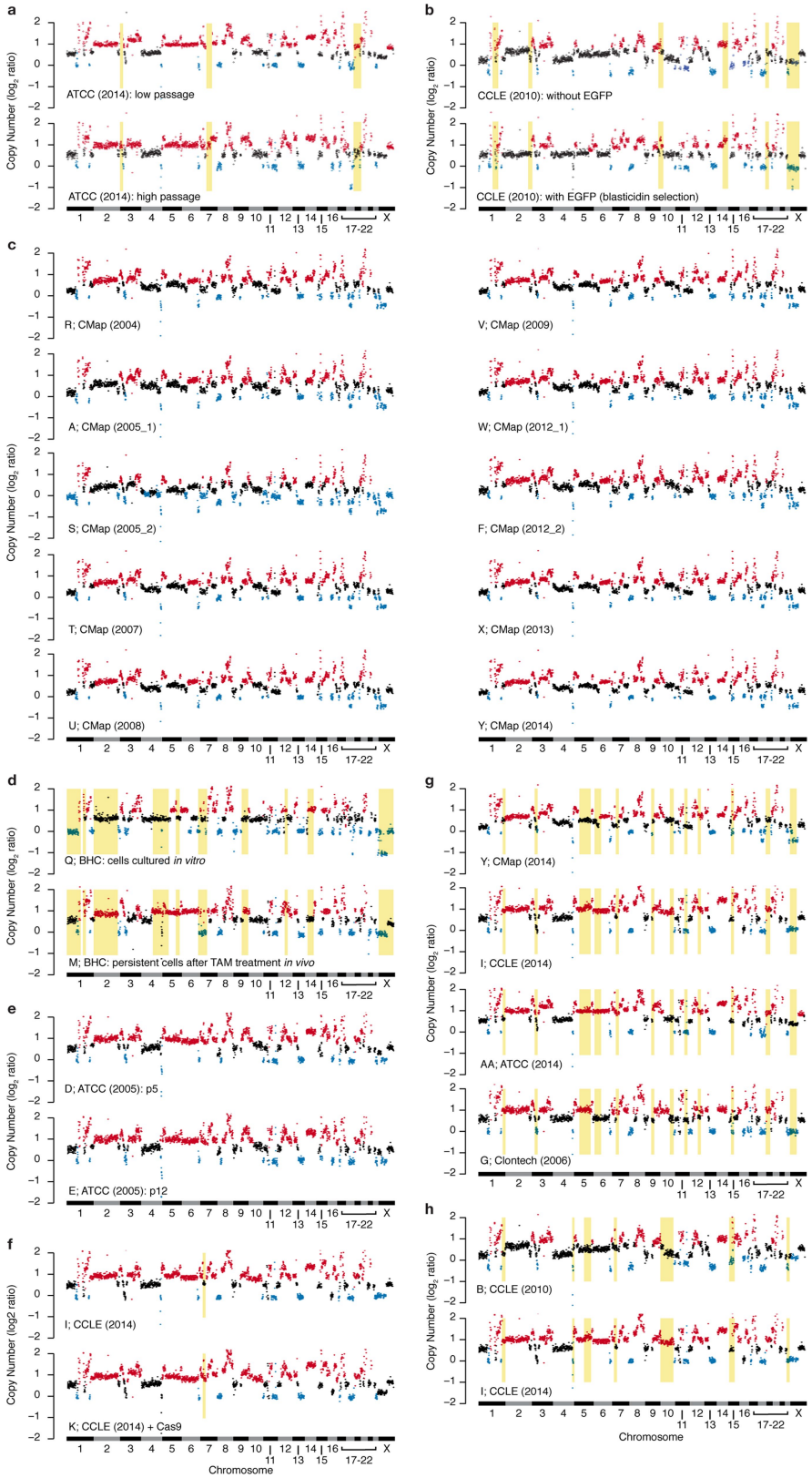
Strain_ID	Origin	Year	Passage	Manipulations	Remarks
A549-A	ATCC/Meyerson				
A549-B	GPP	2014		Cas9-expressing	Derived directly from A549-I
A549-C	CCLE/Hahn	2017	Early passage		Parental of A549-D
A549-D	CCLE/Hahn				Parental of A549-F/G/H
A549-F	CCLE/Hahn			sgRNA against Chr26 (intergenic)	Derived directly from A549-D
A549-G	CCLE/Hahn			sgRNA against TRIB (intergenic)	Derived directly from A549-D
A549-H	CCLE/Hahn			pLX313_Renilla-expressing	Derived directly from A549-D
A549-I	GPP				Parental of A549-B/J
A549-J	GPP			Cas9-expressing	Derived directly from A549-I
A549-K	ATCC/Meyerson				
A549-L	ATCC/Meyerson			DX-HPRT1 (inducible degradation of HPRT1)	
A549-M	ATCC/Amon	2016	Early passage		Parental of MCF7-N/O
A549-N	ATCC/Amon	2016		GFP reporter	Derived directly from A549-M
A549-O	ATCC/Amon	2016	Late passage	Extensive passaging	Derived directly from A549-M
A549-P	Connectivity Map	2009			Parental of A549-Q/R
A549-Q	Connectivity Map	2011		Freeze & thaw	
A549-R	Connectivity Map	2013		Freeze & thaw	
A549-V	ATCC/Hahn	2013	p+11	Cas9-expressing	Derived directly from A549-W
A549-W	ATCC/Hahn	2013	p+10		Parental of A549-V
A549-X	CCLE/PRISM				Parental of A549-Y/Z
A549-Y	CCLE/PRISM			DNA-barcoded (PRISM)	Derived directly from A549-X
A549-Z	CCLE/PRISM			DNA-barcoded (PRISM)	Derived directly from A549-X
A549-AB	ATCC/Brugge				

Extended Data Fig. 2 | Schematic representation of the MCF7 and A549 strains included in the current study. a, MCF7 strains included in this study; their origins (columns), years of acquisition (rows), manipulations (colours) and progeny relationships (lines) are shown. **b**, A table corresponding to **a**. **c**, A549 strains included in this study, their origins (columns), years of acquisition (rows), manipulations (colours) and progeny relationships (lines) are shown. **d**, A table corresponding to **c**.

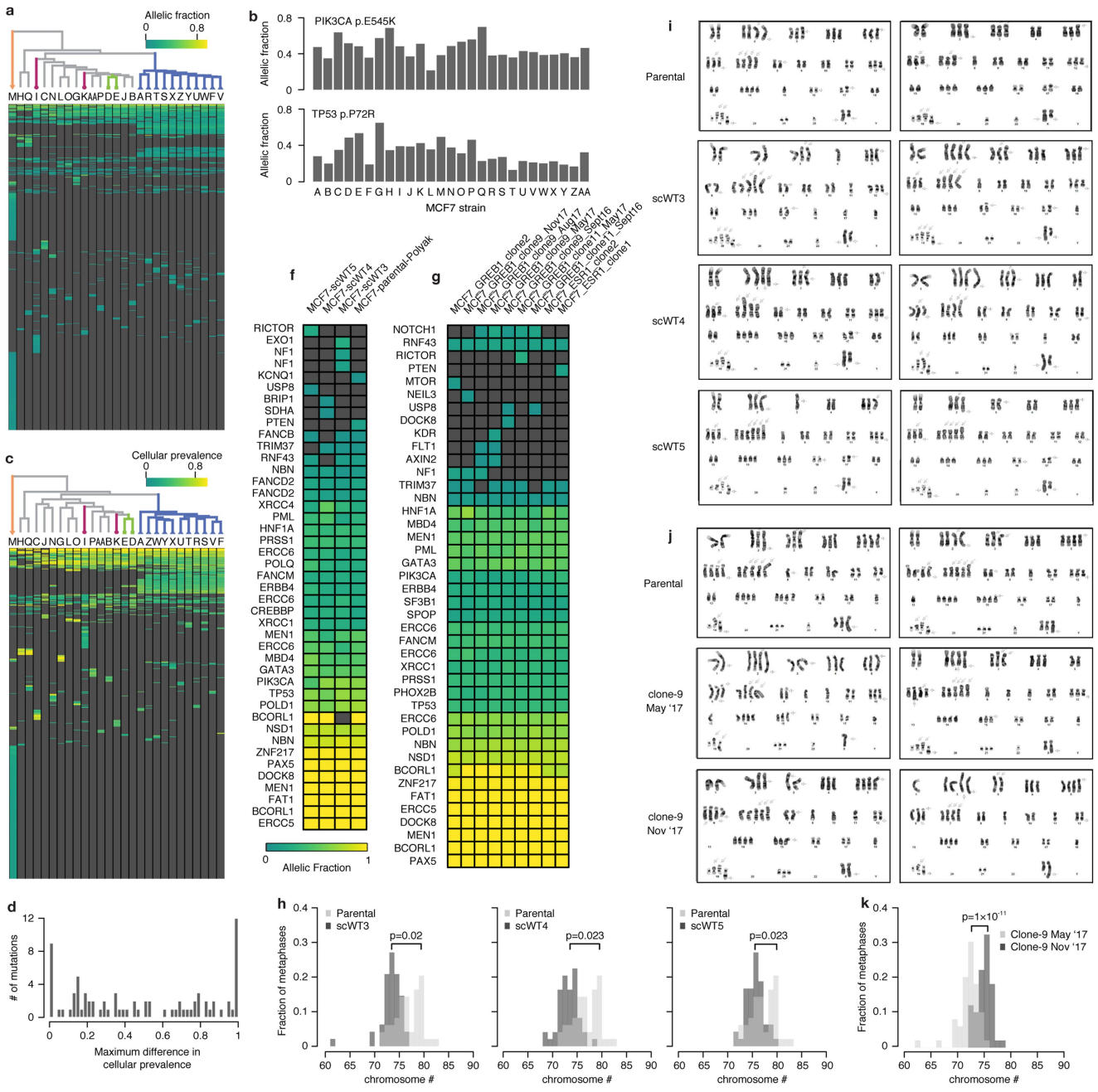


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Genetic variation across 27 MCF7 strains. **a**, Variation in the copy number status of nine selected genes across 27 MCF7 strains. Red, copy number gains; blue, copy number losses. Thresholds for relative gains and losses were set at 0.1 and −0.1, respectively. **b**, Western blots of the relative protein expression levels of ERα across strains. The expression of β-actin was used for normalization. For gel source data, see Supplementary Fig. 1. The experiment was repeated twice with similar results. **c**, Quantification of the relative expression of ERα. Strains Q and M were excluded from the comparison. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile; circles, all data points. One-tailed *t*-test. **d**, The allelic fractions of non-silent mutations in seven selected genes across 27 MCF7 strains. **e**, The number of non-silent point mutations (SNVs) across the 27 MCF7 strains. **f**, The number of COSMIC non-silent point mutations shared by each number of MCF7 strains. **g**, Top, unsupervised hierarchical clustering of 27 MCF7 strains, based on all of their SNVs. Groups of strains expected to cluster together based on their evolutionary history are highlighted, as in Fig. 1. Bottom, a corresponding heat map, showing the mutation status of all mutations across the 27 MCF7 strains. Mutations that were identified in only a subset of the strains that were detected in above 5% of the reads (AF > 0.05) are shown. Yellow, presence of a mutation; grey, absence of a mutation. **h**, The number of large (>15-bp) indels and rearrangements across the 27 MCF7 strains. Grey, indels; black, rearrangements. **i**, The recurrence of structural variants in each of the 42 (out of 60) genes for which at least one event was detected. **j**, The number of structural variants shared by each number of MCF7 strains. Note that this analysis is limited to the 60 genes listed in Supplementary Table 2.



Extended Data Fig. 4 | Comparison of CNA landscapes between MCF7 strains. a, CNA landscapes of a pair of MCF7 strains separated from each other by extensive passaging. **b**, CNA landscapes of a pair of of MCF7 strains separated from each other by a genetic manipulation (introduction of a GFP reporter). **c**, CNA landscapes of 10 MCF7 strains separated by multiple freeze–thaw cycles, with little passaging in between. **d**, CNA landscapes of a pair of MCF7 strains that were either cultured in vitro (top) or cultured in vivo and treated with tamoxifen (bottom). **e**, CNA landscapes of a pair of MCF7 strains separated from each other by seven passages. **f**, CNA landscapes of a pair of MCF7 strains before (top) and after (bottom) the introduction of Cas9. **g**, CNA landscapes of a pair of MCF7 strains obtained from four different sources. **h**, CNA landscapes of a pair of MCF7 strains separated from each other by extensive passaging. Data points represent 1-Mb bins throughout the genome. Red, gains; blue, losses; black, normal copy numbers; yellow, differential CNAs between the compared strains.

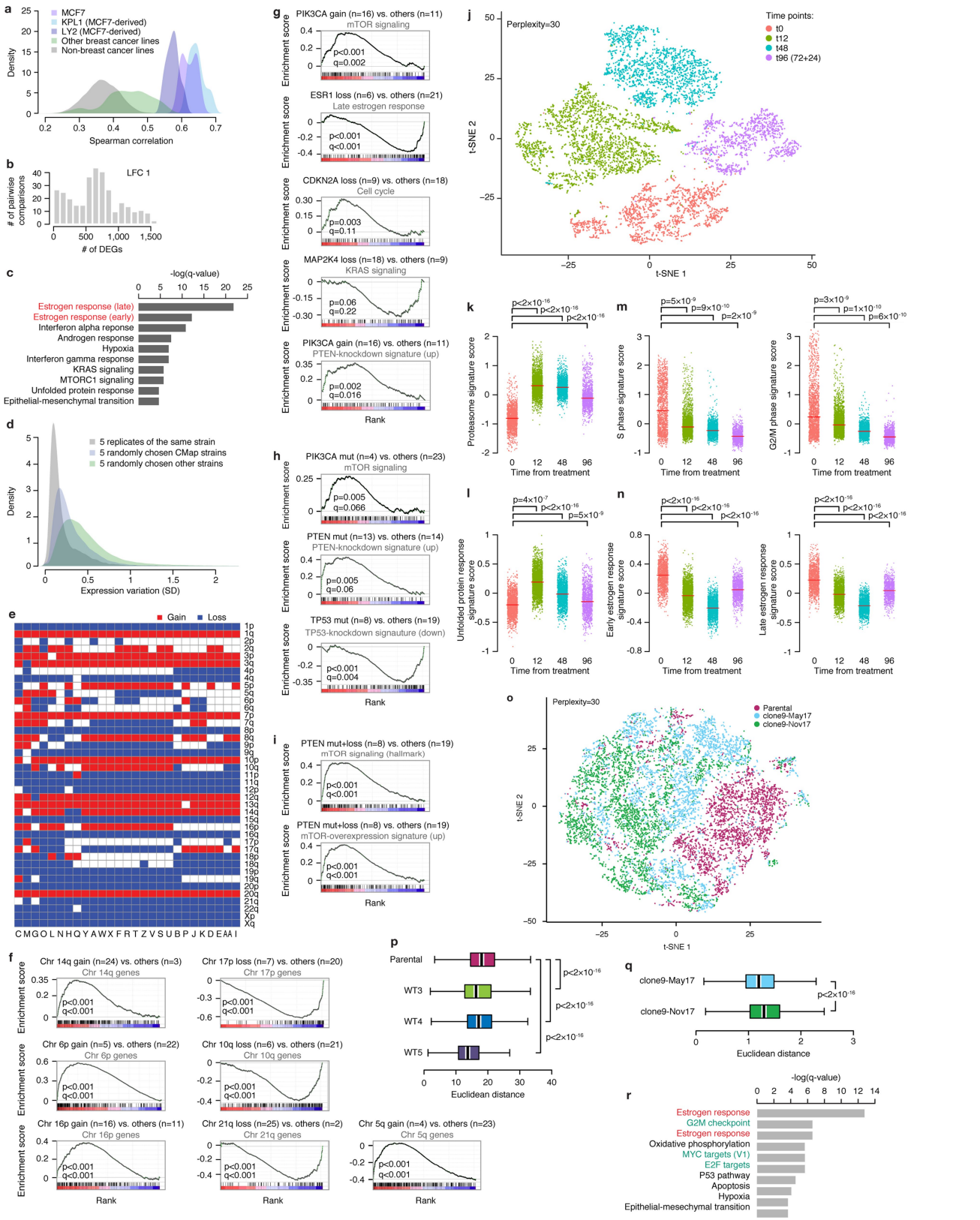


e

Strain_ID	Origin	Year	Clone #	Manipulations	Remarks
MCF7-scWT3	Polyak Lab	2017	3		Derived from MCF7-parental-Polyak
MCF7-scWT4	Polyak Lab	2017	4		Derived from MCF7-parental-Polyak
MCF7-scWT5	Polyak Lab	2017	5		Derived from MCF7-parental-Polyak
MCF7-ESR1_1	Meyerson Lab	2017	1	Luciferase reporter under ESR1 promoter	Derived from MCF7-L
MCF7-ESR1_2	Meyerson Lab	2017	2	Luciferase reporter under ESR1 promoter	Derived from MCF7-L
MCF7-GREB1_2	Meyerson Lab	2017	2	Luciferase reporter under GREB1 promoter	Derived from MCF7-L
MCF7-GREB1_11_Sept16	Meyerson Lab	2016	11	Luciferase reporter under GREB1 promoter	Derived from MCF7-L
MCF7-GREB1_11_May17	Meyerson Lab	2017	11	Luciferase reporter under GREB1 promoter	Derived from MCF7-GREB_11_Sept16
MCF7-GREB1_9_Sept16	Meyerson Lab	2016	9	Luciferase reporter under GREB1 promoter	Derived from MCF7-L
MCF7-GREB1_9_May17	Meyerson Lab	2017	9	Luciferase reporter under GREB1 promoter	Derived from MCF7-GREB_9_Sept16
MCF7-GREB1_9_Aug17	Meyerson Lab	2017	9	Luciferase reporter under GREB1 promoter	Derived from MCF7-GREB_9_May17
MCF7-GREB1_9_Nov17	Meyerson Lab	2017	9	Luciferase reporter under GREB1 promote	Derived from MCF7-GREB_9_Aug17

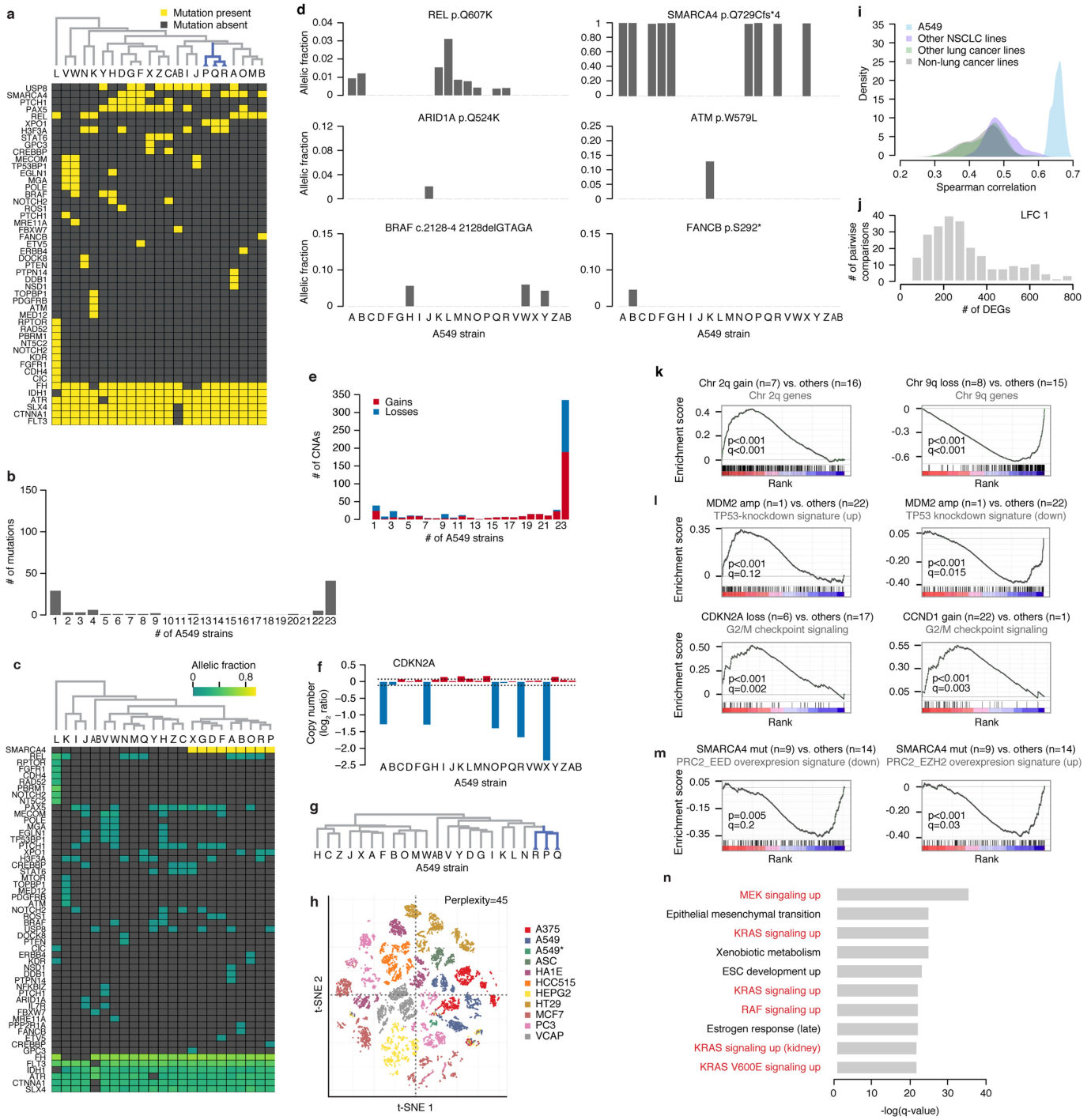
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Characterization of the variation in SNV allelic fraction and cellular prevalence across 27 MCF7 strains and their single-cell-derived clones. a, Top, unsupervised hierarchical clustering of 27 MCF7 strains, based on the allelic fractions of all their SNVs. Groups of strains expected to cluster together based on their evolutionary history are highlighted, as in Fig. 1. Bottom, a corresponding heat map, showing the allelic fractions of all mutations across the 27 MCF7 strains. Mutations that were identified in only a subset of the strains are shown. The presence of a mutation is shown in colour according to its allelic fraction. **b**, The allelic fractions of an activating *PIK3CA* mutation (top) and an inactivating *TP53* mutation (bottom) across strains. **c**, Top, unsupervised hierarchical clustering of 27 MCF7 strains based on their SNV cellular prevalence. Groups of strains expected to cluster together based on their evolutionary history are highlighted, as in Fig. 1. Bottom, a corresponding heat map, showing the cellular prevalence of all mutations across the 27 MCF7 strains. Mutations that were identified in only a subset of the strains are shown. The presence of a mutation is shown in colour according to its cellular prevalence. **d**, The distribution of the maximal differences in cellular prevalence (CP) of non-silent mutations, across 27 MCF7 strains. The peak at maximum $\Delta CP = 1$ represents SNVs that are clonal in at least one strain but are nearly or completely absent in at least one other strain; the peak at maximum $\Delta CP = 0$ represents SNVs that are detected at similar prevalence across all 27 strains; and the peak at maximum $\Delta CP \approx 0.1$ represents a group of SNVs present at $CP \approx 0.1$ only in strain M. **e**, Description of the MCF7 single-cell-derived clones included in this study, including their parental cell line, genetic manipulations and relationship to one another. **f**, A heat map showing the allelic fractions of non-silent mutations in three wild-type single cell-derived MCF7 (scWT3–scWT5) clones and the parental population. The presence of a mutation is shown in colour according to its allelic fraction. **g**, A heat map showing the allelic fractions of non-silent mutations in five genetically manipulated single-cell-derived MCF7 clones. For two of the clones, samples were passaged for a prolonged time and sequenced at multiple time points. The presence of a mutation is shown in colour according to its allelic fraction. **h**, Comparison of the karyotypic variation between parental and single-cell-derived cell populations. Histograms show the distribution of chromosome numbers from the parental (light grey) and single-cell-derived (dark grey) populations. *P* values indicate the significance of the differences between the variations (rather than the means) of the populations using a one-tailed Levene's test ($n = 50$ metaphases per group). **i**, Two representative karyotypes of each sample. Note that all single-cell-derived clones are karyotipically heterogeneous. Marker chromosomes are not shown. Arrows point to partially aberrant chromosomes. Images are representative of 50 metaphases counted per sample. **j**, Two representative karyotypes from two cell populations of the same single-cell-derived clone, separated by six months of culture propagation. Marker chromosomes are not shown. Arrows point to partially aberrant chromosomes. Images are representative of 50 metaphases counted per sample. **k**, Comparison of the karyotypic variation between two cell populations of the same single-cell-derived clone, separated by six months of culture propagation. Histograms show the distribution of chromosome numbers from the early (light grey) and late (dark grey) populations. Per sample, 50 metaphases were counted. The *P* value indicates the significance of the difference between the means of the populations using a two-tailed Wilcoxon rank-sum test.



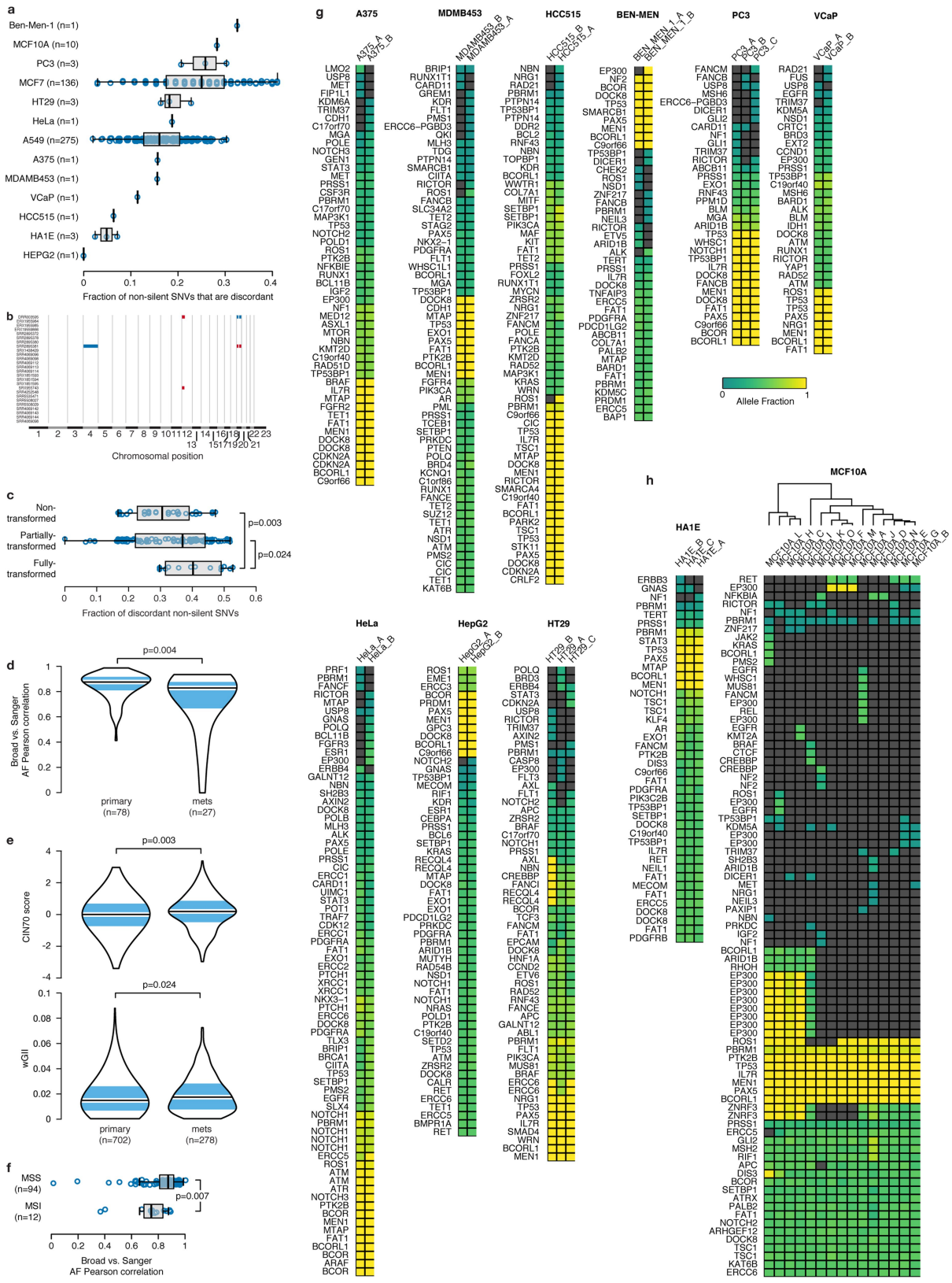
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Transcriptomic variation across 27 MCF7 strains and their single-cell-derived clones. a, Comparison of the L1000-based MCF7 expression profiles to microarray-based expression profiles from CCLE. Histograms show the distributions of the Spearman correlations between the 27 MCF7 strains and either MCF7 (light purple), two MCF7 derivatives (dark purple and blue), other breast cancer cell lines (green) or non-breast cancer cell lines (grey). The comparison is based on the 978 landmark genes directly measured in L1000. **b**, The number of differentially expressed genes identified in all possible pairwise comparisons of MCF7 strains, using a twofold change cutoff. LFC, log fold change; DEGs, differentially expressed genes. **c**, The 10 top hallmark gene sets identified by GSEA to be significantly enriched among the 100 genes that are most differentially expressed across the MCF7 strains. The two gene sets related to oestrogen response are highlighted in red. **d**, Comparison of gene expression variation within and between strains. Histograms show the distributions of gene expression variation within replicates of the same strain (grey), between closely related strains (purple) and between all strains (green). The comparison is based on the 978 landmark genes directly measured in L1000. **e**, Heat map showing the arm-level CNA profiles of 27 MCF7 strains. Red, gains; blue, losses. **f**, GSEA reveals downregulation of the genes on chromosomes 10q, 17q and 21q in strains that have lost copies of these arms, and upregulation of the genes on chromosomes 5q, 6p, 14q and 16p in strains that have gained copies of these arms. **g**, GSEA of the upregulation of mTOR signalling (gene set: hallmark_MTORC1_signalling) and of genes that are upregulated when *PTEN* is knocked down (gene set: PTEN_DN.v2_UP) in strains that have gained *PIK3CA*; downregulation of the oestrogen response signature (gene set: hallmark_oestrogen_response_late) in strains that have lost *ESR1*; cell cycle signature (gene set: KEGG_cell_cycle) in strains that have lost *CDKN2A*; and downregulation of *KRAS* signalling (gene set: hallmark_KRAS_signalling_DN) in strains that have lost *MAP2K4*. **h**, GSEA of the upregulation of mTOR signalling (gene set: hallmark_MTORC1_signalling) in strains with high prevalence of an activating *PIK3CA* mutation; upregulation of genes that are upregulated when *PTEN* is knocked down (gene set: PTEN_DN.v1_UP) in strains that have an inactivating *PTEN* mutation; and downregulation of genes that are downregulated when *TP53* is knocked down (gene set: P53_DN.v1_DOWN) in strains with high cellular prevalence of an inactivating *TP53* mutation. **i**, GSEA reveals upregulation of mTOR signalling (gene sets: MTOR_UP.N4.V1_UP and hallmark_MTORC1_signalling) in strains that have both *PTEN* copy number loss and an inactivating *PTEN* mutation. **j**, A *t*-SNE plot of single-cell RNA-seq data from MCF7-AA cells treated with bortezomib (500 nM) at different time points. Each dot represents a single cell, and cells are coloured by time point. **k**, Comparison of the proteasome gene expression signature across time points. **l**, Comparison of the unfolded protein response gene expression signature across time points. **m**, Comparison of two proliferation gene expression signatures, S (left) and G2M (right), across time points. **n**, Comparison of the early (left) and late (right) response to oestrogen gene expression signatures across time points. Red lines denote mean values. *P* values indicate significance from a one-way ANOVA followed by a Games–Howell post hoc test. $n = 1{,}726$, 2,743, 1,851 and 1,235 cells for t_0, t_{12}, t_{48} and t_{96}, respectively. **o**, A *t*-SNE plot of single-cell RNA-seq data from a parental population and its single-cell-derived clone at two time points. Each dot represents a single cell, and cells are coloured by sample. **p**, Comparison of the transcriptional heterogeneity between a parental MCF7 population and its single-cell-derived clones. $n = 2{,}904$, 2,990, 3,896 and 4,583 cells for parental, scWT3, scWT4 and scWT5, respectively. **q**, Comparison of the transcriptional heterogeneity between two cultures of the same single-cell clone, separated by six months of continuous passaging. $n = 4{,}295$ and 4,116 cells, for clone9-May17 and clone9-Nov17, respectively. Box plots show the Euclidean distance between the cells in each cell population. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile. *P* values indicate significance from a one-way ANOVA followed by a Games–Howell post hoc test. **r**, The 10 top hallmark gene sets identified by GSEA to be significantly enriched among the top differentially expressed genes between the two cultures of clone MCF7_GREB1_9 (May 2017 versus November 2017). The gene sets related to oestrogen response are highlighted in red, and those related to proliferation are highlighted in green.



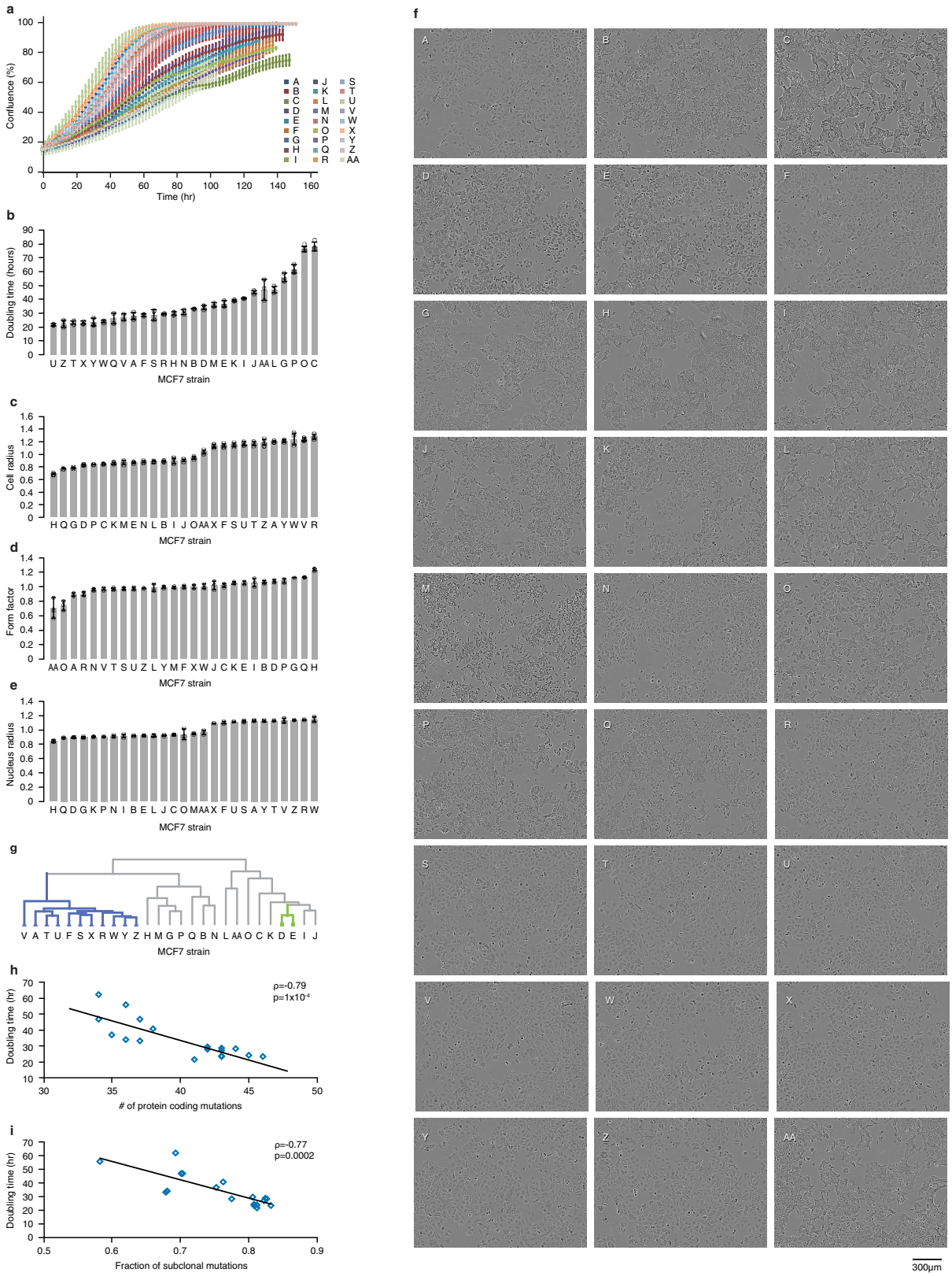
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Extensive genetic and transcriptional variation across 23 strains of A549. a, Top, unsupervised hierarchical clustering of 23 A549 strains, based on their non-silent SNV profiles derived from deep targeted sequencing. Strains expected to cluster together based on their evolutionary history are highlighted in blue. Bottom, a corresponding heat map, showing the mutation status of non-silent mutations across the 23 A549 strains. Mutations that were identified in only a subset of the strains, which were detected in above 5% of the reads ($AF > 0.05$) are shown. The presence of a mutation is shown in yellow, and its absence in grey. **b**, The number of non-silent point mutations shared by each number of A549 strains. **c**, Top, unsupervised hierarchical clustering of 23 A549 strains, based on the allelic fractions of their non-silent SNVs. Bottom, a corresponding heat map, showing the allelic fractions of non-silent mutations across the 23 A549 strains. Mutations that were identified in only a subset of the strains are shown. The presence of a mutation is shown in colour according to its allelic fraction. **d**, The allelic fractions of non-silent mutations in six selected genes across 23 A549 strains. Note the inactivating frameshift mutation in *SMARCA4*, one of the most frequently mutated genes in lung adenocarcinoma[24], which was detected at an allelic fraction of ≈1 in 9 of the strains, but was not detected at all in the other 14 strains. **e**, The number of gene-level CNAs shared by each number of MCF7 strains. Red, copy number gains; blue, copy number losses. **f**, CNA variation in the copy number of *CDKN2A*. Red, copy number gains; blue, copy number losses. Thresholds for relative gains and losses were set at 0.1 and −0.1, respectively. **g**, Unsupervised hierarchical clustering of 23 A549 strains, based on their global gene expression profiles. Strains expected to cluster together based on their evolutionary history are highlighted in blue. **h**, A *t*-SNE plot of L1000-based gene expression profiles from multiple samples of nine cancer cell lines. The asterisk denotes the 23 A549 strains profiled in the current study. **i**, Comparison between the L1000-based A549 expression profiles and the microarray-based expression profiles from CCLE. Histograms show the distributions of the Spearman correlations between the 23 A549 strains and A549 (light blue), other non-small-cell lung cancer cell lines (purple), other lung cancer cell lines (green) or non-lung cancer cell lines (grey). The comparison is based on the 978 landmark genes directly measured in L1000. **j**, The number of differentially expressed genes identified in all possible pairwise comparisons of A549 strains, using a twofold change cutoff. **k**, Arm-level gains are associated with significant upregulation and arm-level losses are associated with significant downregulation of genes transcribed from the aberrant arms. GSEA showing upregulation of the genes on chromosome 2q in strains that have gained a copy of that arm (left), and downregulation of the genes on chromosome 9q in strains that have lost a copy of that arm (right). **l**, Gene-level CNAs are associated with significant dysregulation of the perturbed pathways. GSEA reveals upregulation of the genes that are upregulated, and downregulation of the genes that are downregulated, when *TP53* is knocked down in strains with *MDM2* high-level copy number gain; and upregulation or downregulation of the G2/M cell cycle checkpoint signature in strains with *CDKN2A* copy number loss or *CCND1* copy number gain. **m**, Point mutations are associated with significant dysregulation of the perturbed pathways. For example, GSEA reveals downregulation of two PRC2-related expression signatures in strains with an inactivating *SMARCA4*. **n**, The 10 top gene sets identified by GSEA to be significantly enriched among the 100 genes that are most differentially expressed across the A549 strains. The six gene sets related to *KRAS* signalling are highlighted in red.



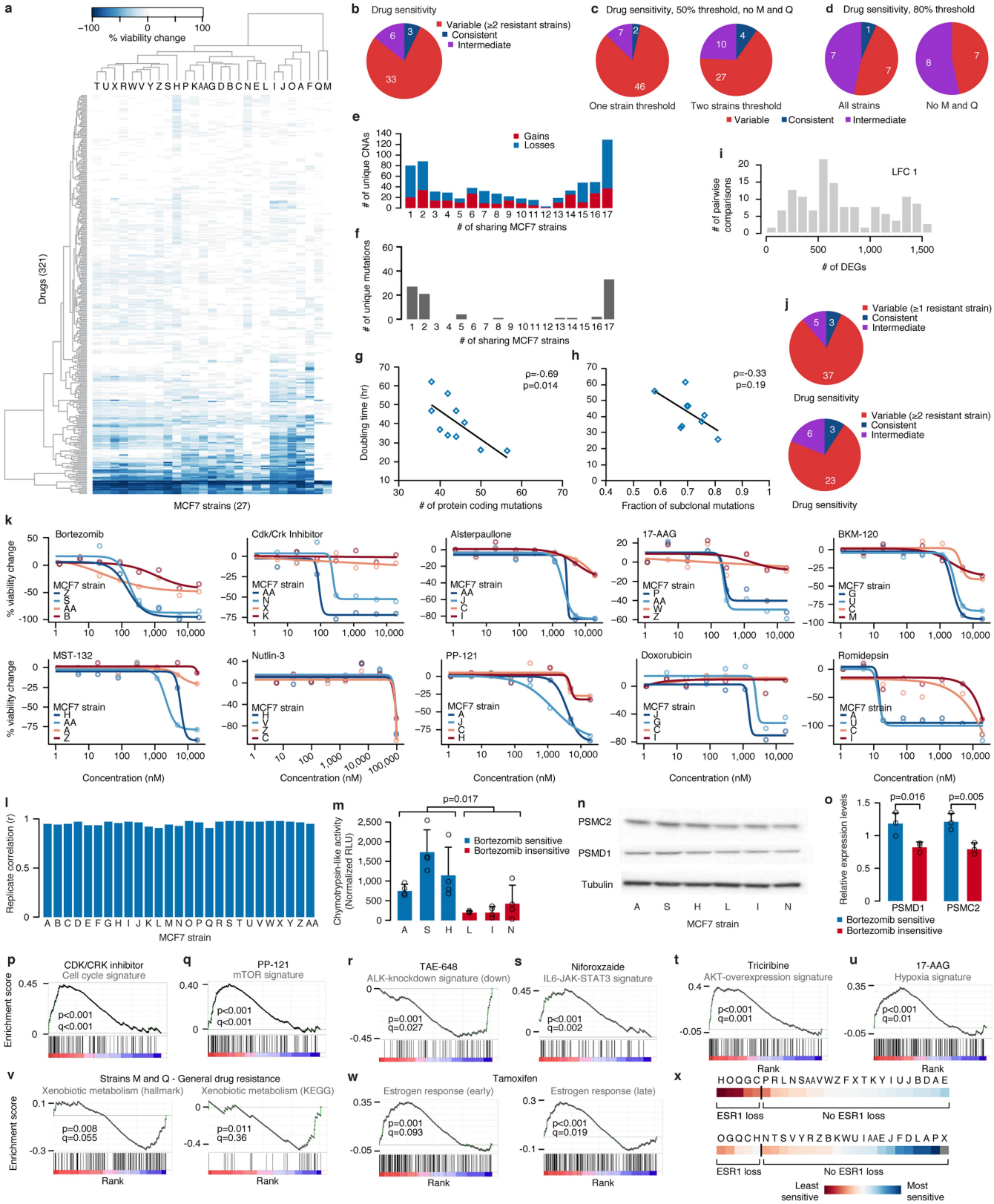
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Genetic variation across multiple strains of additional cancer and non-cancer cell lines. a, The fraction of non-silent SNVs that are discordant between pairs of strains of the same cell line. Data are mean ± s.e.m. *n*, number of strain pairs compared. **b**, Arm-level CNAs arise in RPE1 samples. Plots show CNAs detected by an e-karyotyping analysis of 26 RPE1 samples. Red, gains; blue, losses. **c**, Comparison of variability in non-silent SNVs between non-transformed, partially transformed and fully transformed MCF10A samples. Box plots show the fraction of discordant non-silent SNVs between pairs of samples within each category. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile; circles, all data points. One-tailed Wilcoxon rank-sum test, $n = 28$, 112 and 14 strain pairs, for the non-transformed, partially transformed and the fully transformed groups, respectively. **d**, Comparison of the Broad–Sanger allelic fraction correlations of cell lines derived from primary tumours and those derived from metastases. Bar, median; coloured rectangle, 25th and 75th percentiles; width of the violin indicates frequency at that value. One-tailed Wilcoxon rank-sum test. **e**, Top, comparison of the chromosomal instability (CIN70) gene expression signature score between CCLE lines derived from primary tumours and those derived from metastases. Bottom, comparison of the weighted-genomic integrity index (wGII) between CCLE lines derived from primary tumours and those derived from metastases. Bar, median; coloured rectangle, 25th and 75th percentiles; width of the violin indicates frequency at that value. One-tailed Wilcoxon rank-sum test. **f**, Comparison of the Broad–Sanger allelic fraction correlations of microsatellite-stable cell lines (MSS) and microsatellite-unstable cell lines (MSI). Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile; circles, all data points. One-tailed Wilcoxon rank-sum test. **g**, Heat maps show the allelic fractions of non-silent mutations in multiple strains of cancer cell lines. The presence of a mutation is shown in colour according to its allelic fraction. **h**, Heat maps show the allelic fractions of non-silent mutations in multiple strains of the non-cancer cell lines HA1E and MCF10A. The presence of a mutation is shown in colour according to its allelic fraction. Also shown is an unsupervised hierarchical clustering of the 15 MCF10A strains, which represent different degrees of cellular transformation, based on their non-silent mutation profiles.



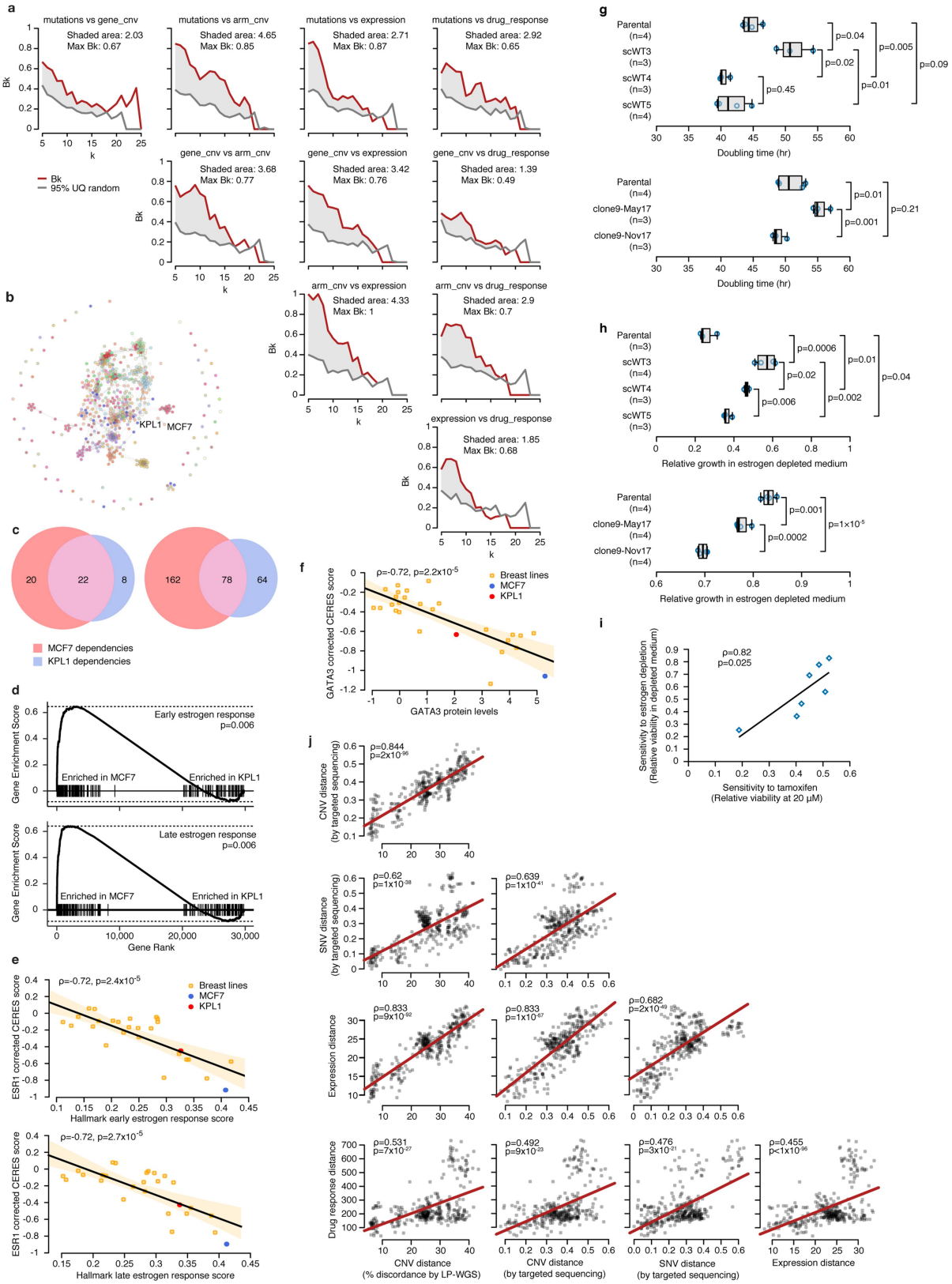
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Characterization of cell proliferation and morphology across 27 MCF7 strains. a, Growth response curves of 27 MCF7 strains, based on microscopy imaging. **b**, Doubling time of the 27 MCF7 strains, as measured by automatic microscopy imaging. **c**, Variation in cellular radius across the 27 MCF7 strains. **d**, Variation in form factor, a measure of circularity, across the 27 MCF7 strains. **e**, Variation in nuclear radius across the 27 MCF7 strains. **a**–**e**, Data are mean ± s.d., circles show individual values; $n = 3$ replicate wells per data point. **f**, Microscopy imaging of the 27 MCF7 strains, showing the morphological differences between them. Scale bar, 300 μm. Images are representative of five replicate wells per strain. **g**, Unsupervised hierarchical clustering of 27 MCF7 strains, based on 1,784 morphological features. **h**, The correlation between proliferation rate (shown as doubling time) and the number of non-silent protein-coding mutations, across 18 naturally occurring MCF7 strains (that is, strains that have not undergone drug selection or genetic manipulation). Spearman's ρ and P values indicate the strength and significance of the correlation, respectively. **i**, The correlation between proliferation rate (shown as doubling time) and the fraction of subclonal mutations, across 18 naturally occurring MCF7 strains. Spearman's ρ and P values indicate the strength and significance of the correlation, respectively.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Characterization of drug-response variation across 27 MCF7 strains. a, Unsupervised hierarchical clustering of 27 MCF7 strains, based on their response to all 321 compounds in the primary screen. Groups of strains expected to cluster together based on their evolutionary history are highlighted, as in Fig. 1. **b**, Pie chart of the classification of the screened compounds based on their differential activity. The response to each active compound was defined as 'consistent' if viability change was $<-50\%$ for all strains, 'variable' if viability change was $<-50\%$ for some strains and $>-20\%$ for other strains, or 'intermediate' if viability change was in between these values. Classification was performed using a two-strain threshold. **c**, Pie charts as in **b** excluding strains Q and M that were generally more drug resistant. Classification was performed using a one-strain or a two-strain threshold (left and right charts, respectively). **d**, Pie charts as in **b** using an activity threshold of viability change $<-80\%$. Classification was performed using a one-strain threshold, either including all strains (left) or excluding strains Q and M (right). **e**, The number of gene-level CNAs shared by each number of MCF7 strains. Red, copy number gains; blue copy number losses. **f**, The number of non-silent point mutations shared by each number of MCF7 strains. The 10 naturally occurring connectivity map strains were averaged and considered as a single sample. **g**, The correlation between proliferation rate (shown as doubling time) and the number of non-silent protein-coding mutations, across naturally occurring MCF7 strains ($n=10$). Spearman's ρ and P values indicate the strength and significance of the correlation, respectively. The 10 naturally occurring CMap strains were averaged and considered as a single sample. **h**, The correlation between proliferation rate (shown as doubling time) and the fraction of subclonal mutations, across naturally occurring MCF7 strains ($n=10$). Spearman's ρ and P values indicate the strength and significance of the correlation, respectively. The 10 naturally occurring CMap strains were averaged and considered as a single sample. **i**, The number of differentially expressed genes identified in all possible pairwise comparisons of MCF7 strains, using a twofold change cutoff. The 10 naturally occurring CMap strains were averaged and considered as a single sample. **j**, Pie charts of the classification of the screened compounds based on their differential activity. The response to each active compound was defined as consistent if viability change was $<-50\%$ for all strains, variable if viability change was $<-50\%$ for some strains and $>-20\%$ for other strains, or intermediate if viability change was in between these values. Classification was performed using a one-strain or a two-strain resistance threshold (top and bottom charts, respectively). The 10 naturally occurring CMap strains were averaged and considered as a single sample. **k**, The dose–response curves for ten compounds are shown. For each compound, eight concentrations were tested in each strain. Two sensitive strains and two insensitive strains are plotted. Each data point represents the mean of two replicates. Nutlin-3, a compound that had no toxicity against any of the strains in the primary screen, was included as negative control. Romidepsin, a compound that killed all strains very efficiently in the primary screen was included as positive control and turned out to be differentially active at lower concentrations. **l**, The Pearson's correlation of the two compound screen replicates across the MCF7 strains. **m**, Strains more sensitive to proteasome inhibitors exhibit higher proteasome activity. The chymotrypsin-like activity of the proteasome was measured in three sensitive and three insensitive strains. Data are mean ± s.d., one-tailed t-test, $n=4$ replicate wells. **n**, Western blots of the relative protein expression levels of the proteasome 19S complex members PSMC2 and PSMD1 in three sensitive and three insensitive strains. The expression of α-tubulin was used for normalization. The experiment was repeated once, with $n=3$ strains per group. For gel source data, see Supplementary Fig. 1. **o**, Quantification of the relative expression of PSMC2 and PSMD1. Data are mean ± s.d., one-tailed t-test, $n=3$ strains per group. **p**, Upregulation of the KEGG cell cycle signature in strains sensitive to the cell cycle inhibitor CDK/CRK inhibitor ($n=3$) compared to insensitive strains ($n=12$). **q**, Upregulation of mTOR signalling in strains sensitive to the PI3K inhibitor PP-121 ($n=11$) compared to insensitive strains ($n=5$). **r**, Downregulation of the genes that are downregulated when *ALK* is knocked down in strains sensitive to the ALK inhibitor TAE-684 ($n=4$) compared to insensitive strains ($n=15$). **s**, Upregulation of IL-6–JAK–STAT3 signalling in strains sensitive to the STAT inhibitor nifuroxazide ($n=9$) compared to insensitive strains ($n=6$). **t**, Upregulation of the genes that are upregulated when *AKT* is overexpressed in strains sensitive to the AKT inhibitor triciribine ($n=2$) compared to insensitive strains ($n=8$). **u**, Upregulation of hypoxia signalling in strains sensitive to the HSP inhibitor 17-AAG ($n=3$) compared to insensitive strains ($n=15$). **v**, Downregulation of xenobiotic metabolism signatures in strains M and Q ($n=2$), which exhibited an increased resistance to most compounds compared to the other strains ($n=25$). **w**, Upregulation of the early and late oestrogen response signatures, in strains most sensitive to the ER inhibitor tamoxifen ($n=5$) compared to the least sensitive strains ($n=5$). **x**, Sensitivity to oestrogen depletion and to tamoxifen is associated with the copy number status of *ESR1*. Heat maps represent the relative viability in oestrogen-depleted medium (top) and in response to tamoxifen (at 16.6 μM; bottom).



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Comparison of genetic-, transcriptomic- and drug-response-based clustering trees, genomic distances and CRISPR dependencies. a, Comparison of clustering trees using the Fowlkes–Mallows approach. The dendrograms were based on SNVs, gene-level CNAs, arm-level CNAs, gene expression profiles and drug response patterns and were all compared to each other. The Fowlkes–Mallows index (Bk) was computed for all potential numbers of clusters (*k* values) ranging from 5 to 26. The red lines indicate the observed Bk values, whereas the grey lines represent the 95% upper quantile of the randomized distribution. The maximum Bk value represents the degree of similarity between the compared pair of dendrograms. The grey shading represents the difference between the observed Bk values and those of the 95% upper quantile of the randomized distribution. **b**, Force-directed layout of screened lines using a similarity matrix determined by the probability of cell lines clustering together in dependency space. Cell lines (nodes) are coloured by lineage. **c**, Left, the overlap of dependencies in KPL1 and MCF7 using corrected CERES scores, with genes showing depletion effects in all cell lines (that is, pan-essential genes) excluded. The threshold for dependency was set as a CERES score <-0.5. Right, overlap in dependency with genes of indeterminate dependency status (CERES scores between -0.4 and -0.6) in either cell line excluded. **d**, A two-sample GSEA of MCF7 and KPL1 against the oestrogen response gene sets ($n=1$ sample per group). Expression of the oestrogen signalling pathway is strongly enriched in MCF7. **e**, The correlation between ESR1 dependency values and the single-sample GSEA enrichment scores of the oestrogen response hallmark gene sets ($n=27$ cell lines). The difference in oestrogen response signalling between MCF7 and KPL1 predicts their differing levels of dependency on ESR1. **f**, The correlation between *GATA3* dependency and GATA3 protein levels (*z*-scored values for reverse-phase protein arrays; $n=27$ cell lines). The difference in GATA3 protein levels between MCF7 and KPL1 predicts their differing levels of dependency on *GATA3*. Spearman's ρ and *P* values indicate the strength and significance of the correlations, respectively. **g**, Top, comparison of proliferation rates between a parental MCF7 population and its single-cell-derived clones. Bottom, comparison of proliferation rates between two cultures of the same single-cell clone, separated by six months of continuous passaging. Box plots show the population doubling time of each sample. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile; circles, all data points. Two-tailed *t*-test; *n*, replicate wells. **h**, Top, comparison of the sensitivity to oestrogen depletion between a parental MCF7 population and its single-cell-derived clones. Bottom, comparison of the sensitivity to oestrogen depletion between two cultures of the same single-cell clone, separated by six months of continuous passaging. Box plots show the relative growth rate in oestrogen-depleted medium. Bar, median; box, 25th and 75th percentiles; whiskers, data within 1.5× IQR of lower or upper quartile; circles, all data points. Two-tailed *t*-test; *n*, replicate wells. **i**, The correlation between sensitivity to tamoxifen (relative viability at 20 μM) and the sensitivity to oestrogen depletion (relative growth rate), across the parental MCF7 populations and their single-cell clones ($n=7$). Spearman's ρ value and *P* values indicate the strength and significance of the correlation, respectively. **j**, Correlation plots between various measures to estimate cell line strains ($n=351$ strain pairs). CNA distances (based on ultra-low-pass whole-genome sequencing or targeted sequencing), SNV distances, gene expression distances and drug response distances were compared to each other. CNA distance based on ultra-low-pass whole-genome DNA-sequencing was determined by the fraction of the genome affected by discordant CNA calls. CNA and SNV distances based on targeted sequencing were determined by Jaccard indices. Gene expression and drug-response distances were determined by Euclidean distances. Spearman's ρ and *P* values indicate the strength and significance of the correlation, respectively.

Extended Data Table 1 | Implications of this study for the use of cell lines in cancer research

Findings	Implications	Recommendations
Given two strains, ~20% of mutations would be observed in only one of them	There is ~10% likelihood that a mutation observed in a strain would not appear in a database of cell line genomic features	• Be cautious when using published datasets of genomic features as "lookup tables"
Prolonged passaging introduces more variation than multiple freeze-thaw cycles	For most cell lines, freezing and thawing is likely to be associated with fewer changes than maintaining in culture	• Keep track of passage number • Use passage-matched controls • For large-scale screens, prepare multiple frozen vials for downstream analyses
Various genomic, transcriptomic and phenotypic assays yield highly similar clustering trees	Simple and inexpensive genome-wide assays can serve as a proxy for diversification	• Use inexpensive genome-wide assays (e.g., LP-WGS) and compare to published references using Cell STRAINER: https://cellstrainer.broadinstitute.org • Exclude strains that show extreme diversification
Genetic manipulations that are considered "neutral" can introduce genetic variation	Cell lines with fluorescent reporters, DNA barcodes or Cas9 expression are not identical to their parental cell lines	• Use efficient infection methods to reduce the bottleneck associated with antibiotic selection • Characterize manipulated strains to ensure they retain hallmark genomic features • In CRISPR screens, correct for copy number effects using the copy number landscape of the screened strain
Genetic and transcriptomic variation may affect drug response	Inconsistencies in drug response studies may be attributed to genetic and transcriptomic variability	• Genetic and transcriptomic distances should be considered when comparing drug response data • Compare drug response data to genomic data from the same strain
	Transcriptional differences between sensitive and resistant strains can elucidate compound mechanism of action	• Use characterized isogenic-like strains to uncover associations between molecular features and drug response
Pre-existing heterogeneity within culture underlies cell line instability	Single cell-derived clones differ from one another genetically, transcriptionally and phenotypically	• Confirm the genomic features of single cell-derived clones • Avoid comparisons between bottlenecked cell populations, whenever possible
	Subtle differences in culture conditions can lead to changes in cell line clonal composition	• Keep culture conditions constant
Heterogeneity keeps emerging in culture due to ongoing genomic instability	Prolonged passaging of single cell-derived clones can lead to their diversification Cell lines with deficient maintenance of genome integrity (e.g., MSI or TP53-mutant) are more prone to genomic evolution	• Re-confirm genomic features of single cell-derived clones following prolonged passaging • Apply these recommendations more stringently to genomically unstable cell lines

A summary of the main findings of this study, their practical implications and recommendations for addressing them.

ARTICLE

https://doi.org/10.1038/s41586-018-0382-x

Creating a functional single-chromosome yeast

Yangyang Shao[1,2], Ning Lu[1,2], Zhenfang Wu[3], Chen Cai[2,3], Shanshan Wang[3], Ling-Li Zhang[2,3], Fan Zhou[4], Shijun Xiao[4], Lin Liu[4], Xiaofei Zeng[4], Huajun Zheng[5], Chen Yang[1], Zhihu Zhao[6], Guoping Zhao[1,5,7,8]*, Jin-Qiu Zhou[3]*, Xiaoli Xue[1]* & Zhongjun Qin[1]*

Eukaryotic genomes are generally organized in multiple chromosomes. Here we have created a functional single-chromosome yeast from a *Saccharomyces cerevisiae* haploid cell containing sixteen linear chromosomes, by successive end-to-end chromosome fusions and centromere deletions. The fusion of sixteen native linear chromosomes into a single chromosome results in marked changes to the global three-dimensional structure of the chromosome due to the loss of all centromere-associated inter-chromosomal interactions, most telomere-associated inter-chromosomal interactions and 67.4% of intra-chromosomal interactions. However, the single-chromosome and wild-type yeast cells have nearly identical transcriptome and similar phenome profiles. The giant single chromosome can support cell life, although this strain shows reduced growth across environments, competitiveness, gamete production and viability. This synthetic biology study demonstrates an approach to exploration of eukaryote evolution with respect to chromosome structure and function.

Almost all known natural eukaryotic species have multiple chromosomes, except for the male ant *Myrmecia pilosula*, which contains only one chromosome[1]. In addition, the number of chromosomes in eukaryotic species varies without a clear association with their biological characteristics. For instance, in mammals, human (*Homo sapiens*) diploid cells have forty-six chromosomes[2], whereas diploid cells of the Indian muntjac (*Muntiacus muntjak*) have the lowest number of chromosomes (six for the female and seven for the male)[3]. In fungi, haploid cells of the budding yeast *Saccharomyces cerevisiae* have sixteen chromosomes and a genome of approximately 12 Mb[4], whereas haploid cells of the fission yeast *Schizosaccharomyces pombe* have only three chromosomes and a genome of approximately 14 Mb[5]. The advantages to a eukaryotic cell of multiple chromosomes instead of a single one are not clear. In this study, we have reorganized the genome of the unicellular eukaryotic model organism *S. cerevisiae*, whose haploid cell contains sixteen chromosomes ranging from 230 to 1,500 kb[4], into one giant chromosome, in order to explore whether a yeast cell with an artificially fused single chromosome can survive and complete a sexual cycle.

Rationale

The creation of a single-chromosome yeast from *S. cerevisiae* BY4742 haploid cells required 15 rounds of chromosome end-to-end fusions, with deletion of 15 centromeres and 30 telomeres (Fig. 1a, Extended Data Table 1). During the fusion process, the following criteria and principles were followed. (1) To generate genetically stable fused chromosomes and avoid the formation of dicentric chromosomes[6], simultaneous deletions of one centromere and two telomeres in each round of fusion were required. We developed a method to precisely fuse two chromosomes by using both the efficient CRISPR–Cas9 cleavage system[7] and the robust homologous recombination activity of yeast (Fig. 1b). (2) The single centromere was intentionally kept roughly in the middle of the final single chromosome to maintain two arms with balanced lengths. (3) The order of chromosome fusion was randomly selected. Our pilot experiments showed that eight pairs of randomly selected chromosomes could all be successfully fused, and the resulting strains grew as robustly as the wild-type strain, indicating that the yeast cells could tolerate random fusion of two chromosomes. (4) The deleted regions of each centromere and telomere were carefully selected to avoid affecting adjacent genes. (5) In addition, the redundant copies of telomere-associated long repetitive sequences (over 2 kb; Extended Data Table 2) located on different chromosomes were deleted to avoid potential homologous recombination at undesired sites.

Chromosome fusion and confirmation

The first chromosome fusion strain, SY0, was constructed by simultaneously removing the telomere and telomere-associated long repetitive sequences of the right arm of chromosome VII (VIIR) and the left arm of chromosome VIII (VIIIL), as well as the centromere element in chromosome VIII (Fig. 1b). Following the same pair-wise fusion strategy, fourteen successive rounds of chromosome fusion were carried out in strain SY0, and finally strain SY14, with all sixteen chromosomes fused into one giant single linear chromosome, was successfully constructed (Extended Data Table 1). For each round of chromosome fusion, the positive rates confirmed by PCR sequencing ranged from 20 to 100% (Extended Data Table 1).

To validate the series of chromosome fusions in strains SY0–SY14, we examined the chromosome numbers using pulsed-field gel electrophoresis (PFGE) under various optimal conditions (Fig. 2a). With the accumulation of chromosome fusions, the DNA bands corresponding to the native chromosomes disappeared accordingly in the lower parts of the gels, and the DNA bands corresponding to the newly fused, larger chromosomes (indicated with red arrowheads) appeared in the upper parts of the gels. The single linear chromosome of SY14 (11.8 Mb) migrated most slowly and remained at the top of the gel (Fig. 2a).

[1]Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]The State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. [4]Frasergen Bioinformatics Co., Ltd, Wuhan, China. [5]Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China. [6]Beijing Institute of Biotechnology, Beijing, China. [7]Department of Microbiology and Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China. [8]State Key Laboratory of Genetic Engineering, Department of Microbiology, School of Life Sciences and Institute of Biomedical Sciences, Fudan University, Shanghai, China. *e-mail: gpzhao@sibs.ac.cn; jqzhou@sibcb.ac.cn; xlxue@sibs.ac.cn; qin@sibs.ac.cn

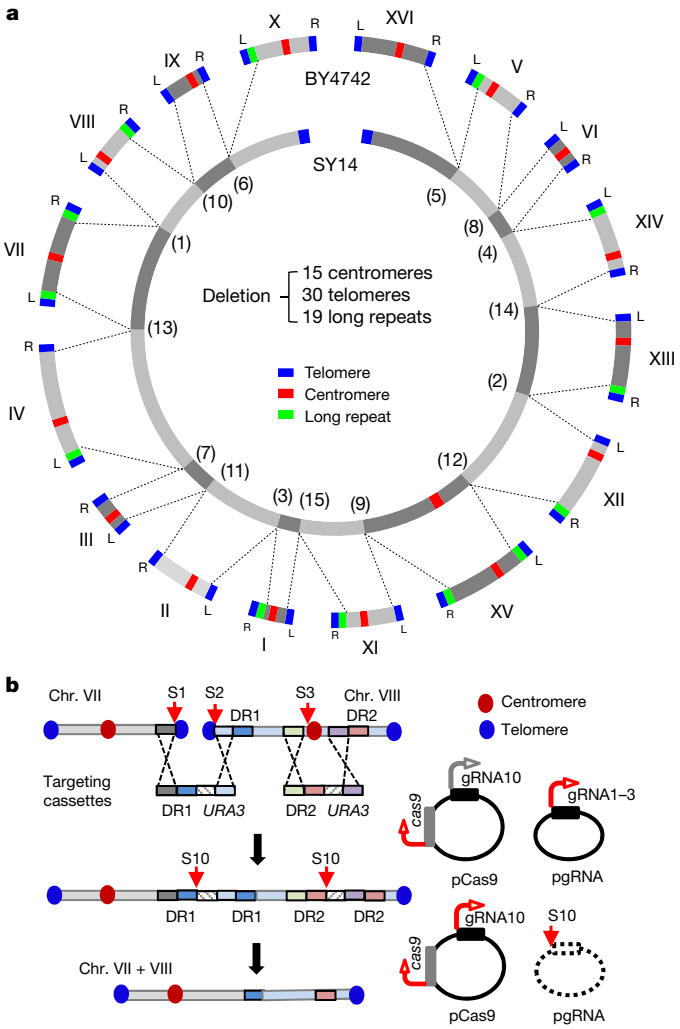


Fig. 1 | Creation of a single-chromosome yeast. a, Sixteen native chromosomes (I–XVI) of BY4742 (wild type) are aligned in the outer ring. The single chromosome of SY14 aligned in the inner ring has undergone fifteen sequential rounds of chromosomal end-to-end fusions, indicated by dashed lines. **b**, CRISPR–Cas9-mediated fusion of chromosomes VII and VIII. Cas9 nuclease cut at the telomere (sites S1 and S2) and centromere (site S3) loci with the guidance of gRNAs 1–3. The broken chromosomes were repaired through homologous recombination with the provided DNA targeting cassettes. The curation of *URA3* and the guide RNA expression plasmid (pgRNA) occurred simultaneously upon galactose induction.

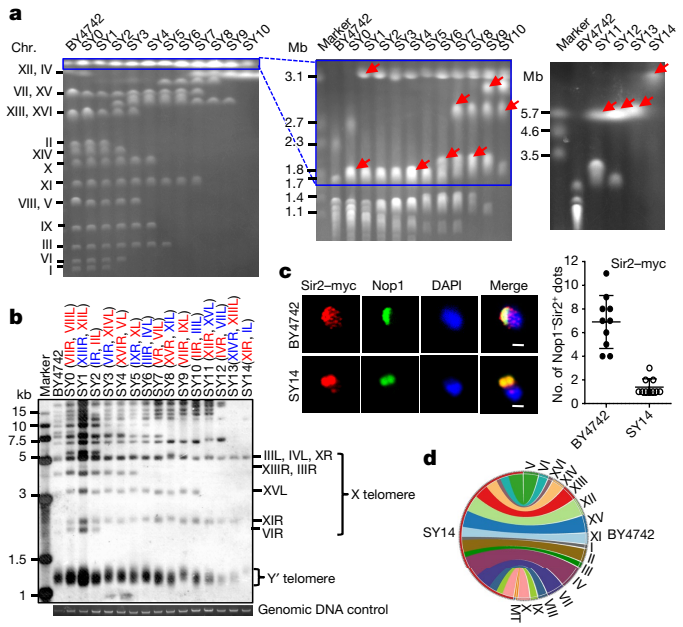


Fig. 2 | Confirmation of chromosome fusion(s) in yeast strains. a, Intact chromosomal DNA analysis by PFGE under different conditions. The red arrows indicate the newly fused chromosomes in each strain. Data shown are representative images of three independent experiment. **b**, The XhoI-digested genomic DNAs from the strains were Southern blotted with a telomere-specific $TG_{1–3}$ probe. The X element- and X-Y′-containing telomeres are indicated in blue and red, respectively, in parentheses. Results are representative of two independent experiments. **c**, The myc-tagged telomere binding protein Sir2 was detected with polyclonal anti-myc antibody and Cy3-conjugated (red) secondary antibody. Nop1, a nucleolar protein, was detected with a monoclonal anti-Nop1 antibody and Alexa 488-conjugated (green) secondary antibody. DNA was stained by DAPI (blue). Scale bar, 1 μm. Data represent mean ± s.d. ($n = 10$ sections per genotype from three independent experiments). **d**, Comparison of the genome sequences of BY4742 (right) and SY14 (left). MT, mitochondrial genome.

The sizes of all chromosomal DNA bands were in agreement with the theoretical calculated sizes (Extended Data Table 1). In addition to PFGE, we performed Southern hybridization with a specific telomeric DNA probe to further confirm the proper chromosome fusions. Following each round of chromosome fusions, the corresponding telomere signals disappeared owing to deletion of these telomere sequences (Fig. 2b, Extended Data Fig. 1). The reduction in telomere numbers in the SY14 strain was visualized by immunostaining of the 13-myc-tagged telomere-binding protein Sir2 using anti-myc antibody[8]. In BY4742 cells, 32 telomeres clustered at 6–8 foci in the nuclei (Fig. 2c), consistent with previous reports[9,10]. Only one or two telomere signals were detected in SY14 cells. Notably, nucleolar structures in the SY14 cells, which could be seen by Nop1 staining (Fig. 2c), remained intact, suggesting that chromosome fusions caused little change in the chromatin region of ribosomal DNA (rDNA) loci.

The genomic DNA of the SY14 strain and its parental strain BY4742 were completely sequenced by a combination of PacBio and Illumina sequencing with 426- and 320-fold coverage, respectively. Their chromosomal nucleotide sequences were de novo assembled into 1 and 16 contigs, respectively. The approximately 1.5-Mb array of rDNA repeats (approximately 9.1 kb)[11] was difficult to assemble, and only several copies of rDNA repeats were assembled in both the BY4742 and SY14 genomes. The complete nucleotide sequences of the single chromosome of SY14 were compared with the sixteen chromosomes of BY4742, which showed excellent co-linearity (Fig. 2d), confirming that the chromosome orderings and orientations were as designed. Sequence alignments of the BY4742 and SY14 chromosomes revealed that all 61 designed deletions (Extended Data Table 1) had been successfully achieved (Fig. 2d, Extended Data Fig. 2); however, 11 single-nucleotide polymorphisms (SNPs) and 7 insertions and deletions (indels) that arose during chromosome fusions were detected using next-generation sequencing and validated by the Sanger method (Extended Data Table 3).

Chromosomal 3D structures

Previous studies have documented the higher-order folding and spatial architecture of all sixteen chromosomes in *S. cerevisae* cells[12,13]. We carried out chromosome conformation capture (3C)-derived Hi-C assays on BY4742 and SY14 cells and on two intermediate strains—SY6 (containing nine chromsomes: seven fused and two native) and SY13 (containing only two fused large chromosomes) (Extended Data Table 1). The global chromosome interactions and average chromosome 3D architecture of the different fusion cells were analysed and compared with those of BY4742 cells[14,15].

The contact maps of BY4742, SY6, SY13 and SY14 cells clearly showed sixteen, nine, two and one independent, distinct intra-chromosomal interaction large square lattice structures, corresponding to the sixteen, nine, two and one individual chromosomes, respectively

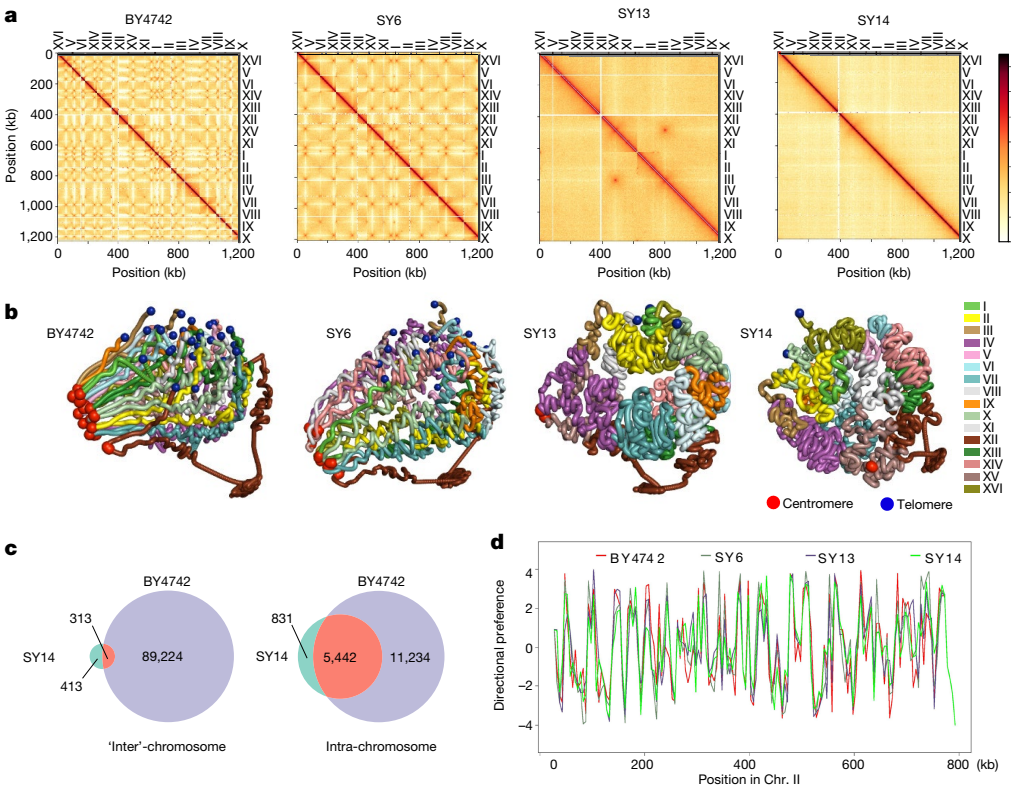


Fig. 3 | Chromosomal interactions and 3D structures of genomes in BY4742, SY6, SY13 and SY14 strains. a, The normalized contact heatmaps of four genomes with 10-kb resolution. Low to high interaction frequencies are depicted by a colour spectrum from light yellow to red. **b**, 3D conformations of the four genomes. **c**, Venn diagram for the numbers of significant ($P < 0.01$, $q < 0.01$) 'inter'- (left) and intra-chromosome (right) interactions. Note that intra- and 'inter'-chromosome here refer to locations in the BY4742 genome. **d**, The directional preferences of chromosome II from four genomes. A paired *t*-test[16] (*y*-axis) assessed the interaction preference of a specific genomic region (genomic coordinates with 5-kb bins shown in *x*-axis) against its upstream (negative *t*-value) or downstream (positive *t*-value) regions.

(Fig. 3a). The *Z*-score difference contact maps showed that the strong centromere–centromere interactions gradually disappeared along with the loss of corresponding centromeres (blue dots), but the interactions among the retained centromeres became much stronger (red dots) (Extended Data Fig. 3a). Consistent with a previous report[12], the sixteen chromosomes of BY4742 cells and the nine chromosomes of SY6 cells showed a typical Rabl configuration[12]: centromeres clustered around the spindle pole body, telomeres clustered with the nuclear envelope, and chromosome arms extending between these two anchoring points (Fig. 3b). Owing to the marked reduction in centromere and telomere numbers, the overall genome structures in SY13 and SY14 cells exhibited a relatively twisted, globular configuration with both the centromeres and telomeres located roughly on the periphery of the whole structure and the two arms of each chromosome much more bent than in BY4742 or SY6 cells, perhaps owing to the nuclear size limitation. It is worth noting that even in the case of the two chromosomes in SY13, the two centromeres and four telomeres were still clustered in roughly opposite positions in the nucleus, similar to the cases of BY4742 and SY6. Notably, the rDNA-repeat loci of all four strains were sequestered from the main structure (Fig. 3b). In particular, the substantial clustering of the flanking sequence of centromeres (red balls) and relative co-localization of the flanking sequence of telomeres (blue balls) in the BY4742 genome gradually disappeared as the chromosome fusion progressed from BY4742 to SY6, SY13 and SY14 (Fig. 3b). Notably, chromosomal fusion in SY6 caused little change to the configurations of the unfused chromosomes, such as chromosome XV, compared to those in BY4742 (Extended Data Fig. 4). However, with the accumulation of chromosome fusions, which resulted in a larger chromosome and a loss of the original centromeres, the 3D structures of chromosomes VI, XVI and X changed from stretched V shapes to more twisted globular shapes (Extended Data Fig. 4).

Almost all (97.8% and 99.7%) of the significant ($P < 0.01$, $q < 0.01$) inter-chromosomal interactions observed in BY4742 were absent in SY13 and SY14, respectively (Fig. 3c), probably owing to the elimination of most of the original centromeres and telomeres (Extended Data Fig. 3a, b). On the other hand, chromosome fusions brought two chromosomes that were distal from each other in BY4742 cells into close proximity, resulting in new significant ($P < 0.01$, $q < 0.01$) inter-chromosomal interactions; for example, the interaction between chromosomes XV and IV in SY13 cells, and chromosomes XV and XII in SY14 cells (Extended Data Fig. 3c). There were ten residual interactions of the single chromosome XV centromere region and chromosome II in all four strains, but the 3D structures of chromosomes XV and II did not show any possible spatial interactions between their centromere regions (Extended Data Fig. 3d). Unlike inter-chromosomal interactions, only 67.4% of total intra-chromosomal interactions were lost in the SY14 genome (Fig. 3c). In fact, the global direction preference[16], which quantifies the preference of a specific genomic region against its upstream or downstream interaction, was similar among BY4742, SY6, SY13 and SY14 cells for each chromosome (Fig. 3d, Extended Data Fig. 5), and the correlation coefficient was 0.90 ($P < 2.2 \times 10^{-16}$). This result strongly indicated that the local chromatin interactions of all the four strains, at least at the level of gene loci (as shown by the bin = 5 kb direction preference plot), were very similar.

Transcriptome and phenome analysis

The transcriptome profiles of the BY4742 and SY14 strains were analysed to evaluate the effects of changes in chromosome interactions and structure on global gene expression. Unexpectedly, the transcriptome of SY14 cells was nearly identical to that of BY4742 cells (Fig. 4a). Only 28 genes were differentially expressed in SY14 compared to BY4742 cells (Fig. 4b, Extended Data Table 4), accounting for 0.5% of the 5,815 protein-coding genes[4,17]. Fusion of all sixteen chromosomes into one repositioned the original telomere-adjacent genes to loci distal from telomeres, which presumably resulted in loss of the telomere position effect (TPE) and de-repression of these genes[18,19]. Accordingly, seven genes (*YFR057W*, *MAL11*, *THI5*, *YOL163W*, *YOL162W*, *SEO1* and *VTH1*) adjacent to the corresponding deleted telomeres were upregulated in SY14 cells. Notably, five genes (*ERR2*, *HSP32*, *FEX2*, *YPL277C* and *MPH3*) near the retained telomeres (XVI-L and X-R) in SY14 cells were downregulated, indicating an increase in the TPE. As about half of subtelomeric genes were deleted during chromosome fusions, the number of TPE-affected genes is likely to be underestimated in

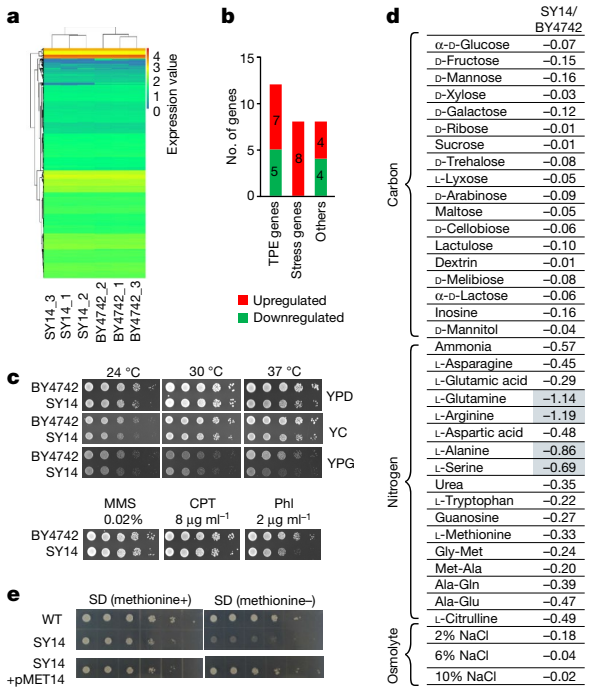


Fig. 4 | Transcriptome and phenome analyses. a, Heatmap of the transcriptome profiles of BY4742 and SY14 cells. The Pearson correlations ($n = 3$) are greater than 0.98 within each group and greater than 0.97 between different groups. **b**, Classification of differentially expressed genes, defined as those with $\log_2$(fold change) ≥ 1 and $P < 0.001$ in SY14 compared to wild-type cells. **c**, Fitness analysis of SY14 cells under various growth conditions. Representative results of two independent experiments. **d**, Growth comparison of BY4742 and SY14 cells under various conditions. The mean area of growth kinetics of SY14 cells from two independent experiments was normalized to those of BY4742 cells, and the numerical value of its logarithm base 2 is shown. The grey shaded negative values indicate greater than 50% growth reduction in SY14 cells. **e**, Restored growth of SY14 cells on medium without methionine by complementation of a functional *MET14* gene. SD, synthetic dextrose. Data are representative of three independent experiments.

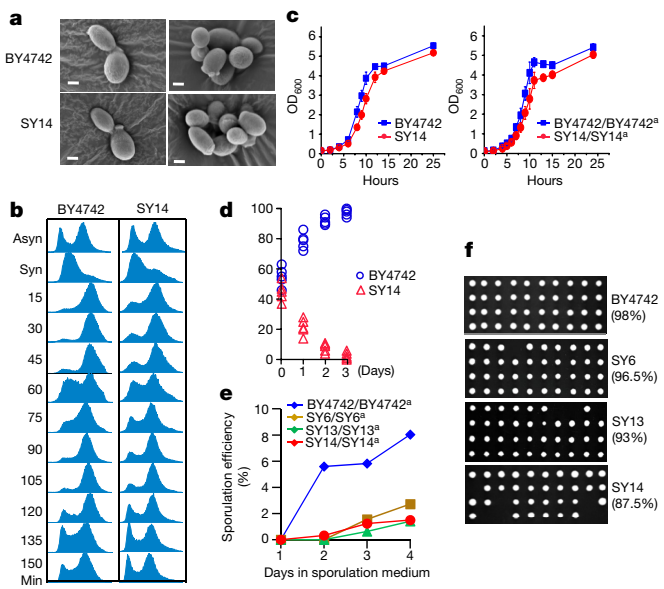


Fig. 5 | Sporulation and competition fitness. a, Scanning electron micrographs of BY4742 and SY14 cells. Scale bar, 1 μm. Representative images of three independent experiments. **b**, Cell cycle analysis. The yeast cells were synchronized with hydroxyurea and the progression of the cell cycle was analysed by flow cytometry. Data are representative of three independent experiments. **c**, Growth curves of the SY14 haploid (left) and diploid (right) strains (mean ± s.e.m.). Three biological replicates were assayed. **d**, Growth competition of BY4742 and SY14 haploid cells. Data from five biological replicates. **e**, Sporulation efficiency of diploid cells with different chromosome numbers. For each genotype, two independent diploid colonies were examined, and the results shared the same trend. **f**, Spore viability of tetrads. Spores of ten tetrads are shown. The spore viability was calculated from 60 tetrads for each strain. Data are representative of two independent experiments.

SY14 cells. Notably, eight genes involved in stress responses (especially DNA replication) were upregulated in SY14, suggesting that a giant single chromosome might introduce a new burden for chromosomal replication.

The SY14 cells demonstrated a slight reduction in growth fitness on complete media such as yeast extract peptone dextrose (YPD), yeast complete (YC) or YPG (similar to YPD but with glycerol as a carbon source), and an increased sensitivity to the genotoxic chemical phleomycin (Phl), but not methyl methanesulfonate (MMS) or camptothecin (CPT) (Fig. 4c). Phenotype microarray analysis showed that SY14 and BY4742 cells had comparable growth under conditions including different carbon sources and osmolytes, but SY14 cells showed a modest growth reduction under some nitrogen sources (Fig. 4d). We found that the expression of the *MET14* gene, which encodes an adenylyl-sulfate kinase, was reduced in SY14 cells comparing to the wild type ($\log_2$ fold-change $= -1.56$, $P = 5.97 \times 10^{-5}$; Extended Data Table 4). When a plasmid-borne *MET14* gene was introduced into SY14 cells, their growth on medium without methionine was restored (Fig. 4e), suggesting that deletion of the chromosome XI centromere accidently damaged the centromere-proximal promoter of *MET14*.

The size and shape of BY4742 and SY14 cells were similar (Fig. 5a). The pattern of cell cycle progression of SY14 cells resembled that of BY4742 cells (Fig. 5b), but the SY14 strain showed a slightly reduced growth rate (Fig. 5c). To evaluate whether the single-chromosome yeast could compete with the multi-chromosome yeast, we co-cultured SY14 and BY4742 cells and monitored their growth. As co-culture time increased, the number of SY14 haploid cells dropped rapidly, while the BY4742 haploid cells dominated the populations (Fig. 5d), suggesting that the competition fitness of the single-chromosome yeast is lower.

Meiosis and spore viability

Organisms that reproduce sexually are thought to have advantages over organisms that reproduce asexually. We evaluated the ability of the single-chromosome haploid cells to mate and form diploid cells and reproduce sexually. We constructed strains BY4742^a and SY14^a, in which the Matα cassette was replaced with a Mata cassette[20]. Haploid SY14 and SY14^a cells were able to mate to produce diploid cells, similarly to the parental strains. But the SY14/SY14^a cells displayed a slightly reduced growth rate (Fig. 5c). In addition, we noticed that two out of six colonies of SY14/SY14^a diploid cells could not maintain a correct diploid chromosome number upon mitotic division. Moreover, the SY14/SY14^a cells displayed weak competitiveness when co-cultured with BY4742/BY4742^a cells, and we observed the emergence of cells that contained genomes from both diploid cells (Extended Data Fig. 6a–c), indicating fusion of SY14/SY14^a and BY4742/BY4742^a cells.

The SY14/SY14^a cells were able to undergo meiosis to produce viable spores, but with reduced gamete production (Fig. 5e). In addition, the spore viability for SY14/SY14^a cells was 87.5%, which was lower than the 98% observed for BY4742/BY4742^a cells (Fig. 5f). The diploid cells of intermediate strains SY6/SY6^a and SY13/SY13^a displayed 96.5% and 93% spore viability, respectively (Fig. 5f), suggesting that spore viability is decreased as the number of chromosome fusions increases.

Discussion

Recently, synthetic biology has made great advances in the design and synthesis of individual chromosomes in the eukaryote *S. cerevisiae*. The synthesized cells largely resemble the wild-type cells[13,21], implying that this organism can tolerate large-scale genome engineering

well. In this study, we created a biologically functional *S. cerevisiae* (SY14) with a single giant chromosome by successive fusion of sixteen native chromosomes, representing, to our knowledge, the first example of a eukaryote with a single linear chromosome created in the laboratory.

Previous studies have suggested that the localization of a chromosome in the nucleus and inter-chromosome interactions affect gene expression[22,23]. In our study, chromosomal fusions involving sixteen chromosomes result in a loss of the majority of the inter-chromosomal interactions seen in parental cells, leading to marked changes in the overall chromosomal 3D structure. However, the global directional preferences at the level of gene loci (5-kb intervals) are largely retained in the SY14 cells. Accordingly, the transcriptome of the single-chromosome SY14 cells is nearly identical to that of the parental BY4742 cells. These observations demonstrate that inter-chromosomal interactions have a negligible effect on global gene transcription in yeast.

It was unexpected that the single point centromere in *S. cerevisiae*, which is only 125-bp long[24], can support the segregation and partition of the 11.8-Mb chromosome, which is eight times larger than the longest native chromosome. Nevertheless, several genes involved in the stress response, especially DNA replication stress, are upregulated in the single-chromosome yeast, consistent with the reported study of an increase in replication-induced topological stress with chromosome length in *S. cerevisiae*[25]. The tendency of SY14/SY14^{a} diploid cells to form polyploidy also suggests a functional defect of chromosome segregation in single-chromosome yeast; this is likely to cause the reduction in gamete production and viability in meiosis. Consistently, both the haploid and diploid cells of the single-chromosome yeast are disadvantaged when competing with wild-type cells. Therefore, the deleterious functional impact of a single giant chromosome, which may be due to chromosome replication and segregation, might explain why eukaryotic genomes are organized into multiple chromosomes. In fact, *S. cerevisiae* and its wild relative species have all maintained sixteen chromosomes across 10–20 million years of evolution, although their chromosome structures are not identical[26]. In an accompanying paper, another group created a two-chromosome budding yeast[27]. Their results are consistent with ours in that the chromosome fusions have minimal effects on cell growth and the transcriptome.

This study provides an alternative approach for studying the evolution of eukaryotes with respect to their chromosome structure and function. The series of strains (SY0–SY14) with successive fusions of sixteen chromosomes created in this study would be of considerable value for future investigations of telomere biology, centromere and kinetochore biology, meiotic recombination, and the relationship between nuclear organization and function.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0382-x.

Received: 29 September 2017; Accepted: 22 June 2018;
Published online 1 August 2018.

1. Crosland, M. W. & Crozier, R. H. *Myrmecia pilosula*, an ant with only one pair of chromosomes. *Science* **231**, 1278 (1986).
2. Green, E. D., Watson, J. D. & Collins, F. S. Human Genome Project: Twenty-five years of big biology. *Nature* **526**, 29–31 (2015).
3. Lee, J. Y. et al. Simple purification of human chromosomes to homogeneity using muntjac hybrid cells. *Nat. Genet.* **7**, 29–33 (1994).
4. Goffeau, A. et al. Life with 6000 genes. *Science* **274**, 546–567 (1996).
5. Wood, V. et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
6. McClintock, B. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).
7. DiCarlo, J. E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
8. Moretti, P., Freeman, K., Coodly, L. & Shore, D. Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.* **8**, 2257–2269 (1994).
9. Gotta, M. et al. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *J. Cell Biol.* **134**, 1349–1363 (1996).
10. Zhou, J., Zhou, B. O., Lenzmeier, B. A. & Zhou, J. Q. Histone deacetylase Rpd3 antagonizes Sir2-dependent silent chromatin propagation. *Nucleic Acids Res.* **37**, 3699–3713 (2009).
11. Zhang, W. et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355**, eaaf3981 (2017).
12. Duan, Z. et al. A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
13. Mercy, G. et al. 3D organization of synthetic and scrambled chromosomes. *Science* **355**, eaaf4597 (2017).
14. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
15. van Berkum, N. L. et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* **39**, 1869 (2010).
16. Le, T. B., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* **342**, 731–734 (2013).
17. Engel, S. R. et al. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda)* **4**, 389–398 (2014).
18. Wellinger, R. J. & Zakian, V. A. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics* **191**, 1073–1105 (2012).
19. Mitchell, L. A. & Boeke, J. D. Circular permutation of a synthetic eukaryotic chromosome with the telomerator. *Proc. Natl Acad. Sci. USA* **111**, 17003–17010 (2014).
20. Xie, Z. X. et al. Rapid and efficient CRISPR/Cas9-based mating-type switching of *Saccharomyces cerevisiae*. *G3 (Bethesda)* **8**, 173–183 (2018).
21. Richardson, S. M. et al. Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).
22. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
23. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
24. Verdaasdonk, J. S. & Bloom, K. Centromeres: unique chromatin structures that drive chromosome segregation. *Nat. Rev. Mol. Cell Biol.* **12**, 320–332 (2011).
25. Kegel, A. et al. Chromosome length influences replication-induced topological stress. *Nature* **471**, 392–396 (2011).
26. Yue, J. X. et al. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* **49**, 913–924 (2017).
27. Luo, J., Sun, X., Cormack, B. P. & Boeke, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* https://doi.org/10.1038/s41586-018-0374-x (2018).

Acknowledgements We thank X. Gao, W. Zhao, and J. Li for technical help. This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB19000000), the National Natural Science Foundation of China (31421061, 31770099, 31370120, 31230040), the National Key Basic Research Program of China (973 Program) (2011CBA00801, 2012CB721102), and the National Key Research and Development Program of China (2016YFA0500701).

Reviewer information *Nature* thanks G. Liti, K. Wolfe and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Z.Q. and X.X. designed and analysed all experiments. J.-Q.Z. and G.Z. contributed to the experiment designs and data evaluation. Y.S. constructed the single-chromosome yeast and conducted the scanning electron microscopy characterization. N.L. conducted the PFGE confirmation, growth characterization and cell cycle experiments. C.C. and L.-L.Z. conducted meiosis and genotoxin sensitivity experiments. Z.W. and S.W. conducted telomere characterization. S.X., X.Z. and H.Z. performed genome analysis. Z.Z. designed the Hi-C and part of the RNA-seq experiments and data interpretation. F.Z., L.L. and Z.Z. performed chromosome Hi-C data analysis. F.Z. analysed the RNA-seq data. X.X. and C.Y. analysed the phenotype microarray data. X.X. wrote the primary manuscript with contributions from J.-Q.Z., Z.Q., G.Z., Z.Z., C.Y., S.X. and Z.W.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0382-x.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0382-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to G.Z. or J.-Q.Z. or X.X. or Z.Q.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Plasmid constructions. The Cas9 expression plasmid pCas9 was constructed from pMetCas9[28] by replacing the selection marker *MET14* with *LEU2*. The guide RNA expression plasmid (pgRNA) was constructed in three steps: (1) the vector pHIS426 was constructed by Gibson assembly[29] of the *HIS3* gene amplified from *S. cerevisae* S288C genomic DNA and vector backbones amplified from p426-SNR52p-gRNA.CAN1.Y-SUP4t. (2) Each guide RNA expression cassette contains the *SNR52* promoter, a 20-bp target sequence and the structural component of guide RNA, followed by the SUP 3′ flanking sequence. The 20-bp target sequences of guide RNAs 1, 2 and 3 were manually selected from the upstream of any 5′-NGG near the to-be-deleted centromere and telomeres. The guide-RNA expression cassettes were generated by fusion PCR of two segments, the SNR52 promoter and the guide RNA structural component/the SUP 3′ flanking sequence segment, both of which were PCR amplified using p426-SNR52p-gRNA.CAN1.Y-SUP4t (Addgene plasmid ID: 43803) as a template. For each guide RNA expression cassette, a 20-bp RNA target sequence and specific restriction sites were introduced at the 5′ end of PCR primers. (3) The three target cassettes gRNA1, gRNA2 and gRNA3 were digested with pairs of restriction enzymes EcoRI–BamHI, BamHI–NcoI and NcoI–NotI, respectively, and were ligated to an EcoRI–NotI-digested pHIS426.

CRISPR–Cas9 facilitated chromosome fusion. Approximately 1 μg of each DNA targeting cassette (with 50–400 bp homology arms and 200 bp direct repeat (DR) sequences for the curation of selection markers in the second step) and pgRNA were co-transformed in *S. cerevisae* BY4742 (Euroscarf, not tested for mycoplasma) cells harbouring pCas9, which constitutively expressed Cas9, using a standard lithium acetate transformation protocol[30]. The transformation products were plated on the synthetic drop-out medium SC-Ura-His-Leu (omitting uracil, histidine and leucine). The positive colonies verified by PCR sequencing were inoculated and grown in 3 ml of SC-Ura-His-Leu liquid medium to saturation at 30 °C. The cell cultures were then transferred to SC-Leu medium containing 2% galactose and 3% raffinose instead of 2% glucose, with an initial optical density of $OD_{600} = 0.3$. After 16 h, 100 μl of culture was plated and grown on SC-Leu with 1 mg/ml 5-FOA. The curation of selection markers and pgRNA of the positive colonies was verified by PCR analysis and sequencing. The verified single colony was inoculated in SC-Leu medium to start the next round of chromosome end-to-end fusion.

Telomere Southern blot. Telomere Southern blotting and hybridization were performed as previously described[31]. In brief, roughly equal amounts of genomic DNA were digested with XhoI, and separated by electrophoresis on 1.0% agarose gel. The DNA was then transferred to a Hybond-N+ Nylon membrane (GE Healthcare). Probe labelling, hybridization and immunological detection were performed using DIG-High Prime DNA Labelling and the Detection Starter Kit II (Roche). An 81-bp TG_{1-3} DNA fragment was chosen as a telomere-specific probe, and a fragment from chromosome I served as a non-telomeric control probe.

Genome sequencing, assembly and data accessibility. A total of 20 μg of high quality genome DNA was extracted from BY4742 and SY14 cells. A 20 kb SMRT-bell sequencing library (Pacific Biosciences) was constructed using a size selection protocol on the BluePippin (Sage Science). The two SMRT-bell yeast genomic libraries were sequenced using 3 SMRT cells (Pacific Biosciences) with a 10-h moving time window in the PacBio Sequel (Pacific Biosciences) sequencing platform. Primary filtering on polymerase reads was performed using the SMRT analysis package V4.0 (https://www.pacb.com/support/software-downloads/).

To assemble the BY4742 and SY14 genomes, 320× and 426× of PacBio subreads were used, respectively. The genomic sequence data of SY14 and BY4742 genomes were directly assembled into 1 and 16 contigs using CANU[32] (version 1.5) without additional data or scaffolding steps. The assembled genomes have no Ns in their sequences. Owing to the high number of repeats in the rDNA region (100–200 copies[11] of ~9,100 bp unit), we could not assemble all repeat sequences for this highly repeated region. Using longer reads of PacBio sequencing, we have assembled a repeated region longer than reported in the public S288C reference genome (GCF_000146045.2_R64/). The nuclear genomes were revised using pbalign (https://github.com/PacificBiosciences/pbalign, version 0.3.1, algorithm: blasr, hitPolicy: randombest, algorithmOptions:–bestn 1–nCandidates 1–bam, maxDivergence 15.0, minAccuracy 85.0) and arrow (https://github.com/PacificBiosciences/GenomicConsensus, version 2.2.0, minCoverage 15). The mitochondrial genomes of BY4742 and SY14 were assembled using minimap (version 0.2-r124-dirty) and miniasm (version 0.2-r137-dirty)[33]. The revision of mitochondrial genomes of BY4742 and SY14 cells was performed using the same procedure with modified parameters of max Divergence 30.0, min Accuracy 70.0 (pbalign) and min Coverage 30 (arrow). The mitochondrial genomes were cyclized with minimus2 (https://github.com/sanger-pathogens/circlator/wiki/Minimus2-circularization-pipeline, AMOS, version 3.1.0).

The genomes of BY4742 and SY14 cells were aligned using LAST[34] (version 810). The scripts last-split[34] and maf-swap were used to achieve 1-to-1 alignment according to the manual (http://last.cbrc.jp/doc/last-split.html). The alignments between two genomes were extracted from LAST output using a custom Python script and plotted with Circos[35] (version 0.67-7). To correct for possible error of Single Molecule Real-Time (SMRT) sequencing reads (predominantly either deletions or insertions) and to obtain the reliable genomic difference between BY4742 and SY14, we re-sequenced both SY14 and BY4742 whole genomes using next generation sequencing (NGS) Illumina pair-end sequencing, which provided 200× coverage for SY14 and 233× coverage for BY4742 cells. These new sequence data were mapped to the S288C genome to identify the genomic differences between the two strains, which were further validated by Sanger sequencing.

Hi-C library preparation and sequencing. The genomic DNA from exponential phase cells was cross-linked and digested with 200 unit MboI enzyme (NEB) as previously described[36]. The cutting DNA ends were labelled with biotin-14-dCTP (TriLINK) followed by ligation. Purified DNA was sheared to a length of ~400 bp. Point ligation junctions were pulled down with Dynabeads MyOne Streptavidin C1 (Thermofisher). The Hi-C library for Illumina sequencing was prepped with the NEBNext Ultra II DNA library Prep Kit for Illumina (NEB) according to the manufacturers' instructions. Fragments between 400 and 600 bp were paired-end sequenced on a HiSeq2000 platform (Illumina).

Construction of contact map and chromosome 3D model from Hi-C data. Using the ICE software package (version 1f8815d0cc9e)[37], the Hi-C data of BY4742 cells were iteratively mapped to the BY4742 genome, while the Hi-C data of SY6, SY13 and SY14 cells were mapped to their own genomes. Dandling ends and other unusable data were filtered, and the valid pairs were binned into 10-kb non-overlapping genomic intervals to generate contact maps. The contact maps were normalized using an iterative normalization method to eliminate systematic biases. To correct the overall decay of chromatin contacts with genomic distance, the whole genome interactions of BY4742 compared to SY6, SY13 and SY14 cells were transformed into *Z*-score using E. Crane's method[38] (package cworld-dekker from https://github.com/dekkerlab/cworld-dekker/releases, v1.01). Then, we calculated the difference between the *Z*-scores of SY14, SY13 and SY6 to BY4742 to infer the difference between their whole genome interactions. The chromosomal 3D structures of the four strains were inferred using the Pastis (v0.1) method[39] with a multidimensional scaling (MDS) model. The 10-kb contact maps were used to construct the 3D model. The rDNA region was reconstructed by assuming that every bin in rDNA loci was equally in contact with the remainder of the genome.

Calculation of intra-and inter-chromosome interactions. As the correlation efficiency of the two biological replicates for each strain was very high (>0.98) using QuASAR-Rep analysis (from HiFive[40] v1.5.3)[41], we pooled the data from two replicates together for significant interactions. The contacts between 5-kb pairs of intra-chromosome bins of four strains were transferred to Ay's Fit-Hi-C software (v1.0.1)[42] to calculate the corresponding cumulative probability *P* value and false discovery rate (FDR) *q* value[43]. After calculation, the interactions in which both the *P* value and *q* value were less than 0.01 were identified as significant interactions.

The statistical significance and false discovery rate of the inter-chromosome interactions of BY4742, SY6 and SY13 cells were calculated using Xie's method[44]. We then calculated the FDR *q* value, and the interactions in which both the *P* value and *q* value were less than 0.01 were identified as significant interactions.

Comparison of significant interactions. Interactions were considered significant at $P < 0.01$, $q < 0.01$.The coordinates of SY14, SY13 and SY6 genome bins were mapped to those in the BY4742 genome by sequence alignment. For better comparison, we separated the interactions into intra-chromosome and 'inter'-chromosome (between original chromosome regions) and performed pairwise comparison. We defined the sequences of 100 MboI-cutting sites around the centromere as the centromere regions, and the sequences of 100 MboI-cutting sites adjacent to the telomeres as the telomeres regions.

RNA-seq analysis. The BY4742 and SY14 cells from early-log phase ($OD_{600} = 1$) were collected, and their RNA was isolated using the TRIzol (Invitrogen) method[45]. The library preparation followed the standard procedure (BGI). The libraries were sequenced on the Illumina HiSeq 4000 platform using the 150-bp pair-end sequencing strategy. For each sample, 6 Gb of clean data was obtained. The cleaned reads were mapped onto the *S. cerevisiae* S288C reference genome (GCF_000146045.2_R64/) using Bowtie2 (v2.2.2)[46], and the abandon estimation was conducted by RSEM (v1.3.0)[47]. The significant differentially expressed genes (DEGs) were identified by DEseq2 (v1.16.1)[48] with definition of fold change more than 2 and false discovery rate (FDR) <0.001.

Phenotypic microarray analysis. Cells freshly grown on YPAD plates were inoculated into a yeast nutrient supplement mixture (NS × 48-0.48 mM L-histidine HCl, 4.8 mM L-leucine, 2.4 mM L-lysine HCl, and 1.44 mM uracil) and adjusted to a transmittance of 62% T using a Biolog turbidimeter (Biolog). Phenotypic microarray plates PM1–PM10 (Biolog) coated with different nutrients and chemical substrates

were used to incubate three cell types. For PM1-8, Biolog growth medium was prepared using IFY-0 (IFY-0 × 1.2) base inoculating fluid, supplemented with a mixture of 0.02% (v/v) yeast nutrient supplement mixture (NS × 48) and 0.013% (v/v) dye mix D (Biolog). To the growth medium, an extra 100 mM D-glucose had to be added for PM3-8. For PM9 and 10, Biolog growth medium was prepared using 0.67% (w/v) YNB w/o amino acids (Sunrise Science) and 0.12% (w/v) SC amino acid mixture (Sunrise Science) supplemented with a mixture of 0.01% (v/v) yeast nutrient supplement mixture (NS × 48), 0.01% (v/v) dye mix E (Biolog) and 100 mM D-glucose. The final volume of 12 ml was reached using reverse osmosis (RO) sterile water for every phenotypic microarray plates and added at 100 μl/well. Data were recorded photographically at 15 min intervals at 30 °C for 120 h and converted to a value reflecting metabolic activity by the OmniLog software (version 2.3.01).

Cell growth, morphology and cell cycle analysis. Strains BY4742 and SY14 were freshly streaked on plates, and three individual colonies were picked and inoculated in liquid YPAD medium overnight at 30 °C. The cell cultures were harvested and diluted in 25 ml of fresh liquid YPAD medium to a final OD_{600} of ~0.1. The optical density of cells was measured hourly and exponentially growing cells were collected. The samples were prepared for scanning electron microscopy (Zeiss) as described previously[49].

Methods of cell synchronization and cell cycle analysis using flow cytometry (Beckman) were performed as previously described[50]. In brief, the yeast strains were synchronized with 200 mM hydroxyurea for 1.5 h. Then, cells were washed five times with pre-warmed YPAD for release from hydroxyurea. The cells were collected every 15 min, washed, and fixed in 70% ethanol at 4 °C overnight. Cells were then treated with RNase (20 mg/ml, Sigma) at 37 °C for 2–3 h. Samples were stained by PicoGreen (Invitrogen), and analysed by flow cytometry. Approximately, 10^5 cells were analysed for each strain. Data were analysed with Summit 5.2.

Genotoxin sensitivity assay. Single colonies of the tested strains were cultured in YPD medium at 30 °C overnight. The cell numbers were adjusted to about equal for each strain, and tenfold serial dilutions were spotted onto plates with or without the indicated genotoxins. The plates were photographed after incubation at 30 °C for 2 to 3 days. For temperature sensitivity assay, plates were incubated at 24 °C, 30 °C and 37 °C for 2 to 3 days before photography.

Growth competition. The *HIS3* and *URA3* genes were introduced into BY4742 and SY14 chromosomes, respectively. Approximately 1×10^4 exponentially grown cells of BY4742 and SY14 haploid cells were co-inoculated in fresh YPD medium at 30 °C. The co-cultures were diluted to 0.05% daily with fresh YPD. The co-cultures after 0, 1, 2 and 3 days were plated on YPD plates, and 100 colonies grown from these plates were spotted on both SC-His (synthetic complete medium without histidine) and SC-Ura (synthetic complete medium without uracil) plates to calculate the viable cells of BY4742 and SY14, respectively. The growth competition of BY4742/BY4742^a and SY14/SY14^a diploid cells were carried out similarly.

Immunofluorescence. A 13-myc epitope-tag sequence was inserted at the genomic locus of the *SIR2* gene. Indirect immunofluorescence was performed as described previously[10]. Cells were grown in YPD medium overnight to a density of $1–2 \times 10^7$ cells/ml and fixed for 30 min by incubation with 4% formaldehyde. Next, cells were washed with 0.1 M potassium phosphate (pH 6.5) and P solution (1.2 M sorbitol and 1 M K_2PO_4), and re-suspended in P solution. Cells were subsequently treated with 0.1 mg/ml zymolyase (20T, MP Biomedicals) for 10 min, washed with P solution, and spotted on poly-L-lysine pre-treated slides. After rinsing in PBS-T buffer (PBS containing 0.1% Triton X-100 and 1% BSA), slides were incubated overnight at 4 °C with anti-Myc, anti-Rap1 and anti-Nop1 antibody diluted in PBS containing 1% BSA. Slides were then washed with PBS-T buffer and incubated with the appropriate secondary antibodies conjugated to Cy3 or Alexa 488. The DNA fluorescence signal was detected by DAPI (1 μg/ml in phosphate buffered saline (PBS) solution) staining. Slides were mounted with PBS containing 1 mg/ml p-phenylenediamine, 2.5 μM NaOH, and 90% glycerol. Confocal microscopy (Leica) was performed on a Leica TCS SP2 microscope with a 63 × lambda blue objective (oil). Image processing included similar filtration and threshold level standardization for all images.

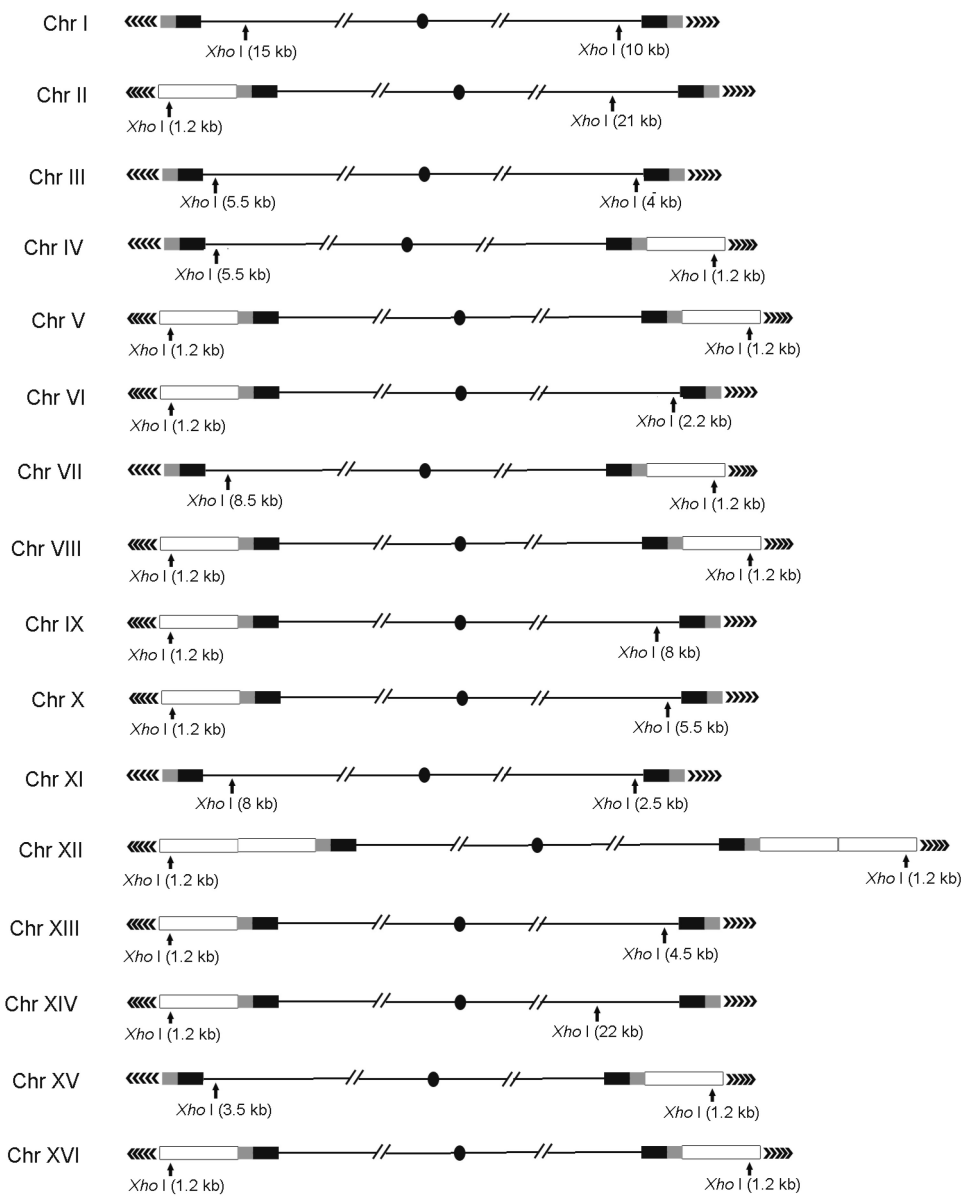
Switching the MATα mating-type of the SY14 strain to MATa to generate the SY14^a strain. Most sequence of the 'MATa' and 'MATα' loci in different mating type yeasts are identical except that a 703 bp region in the MATa locus (5′-cttgatttttgttctttcggggaa…tgtagagtggtttgacgaataatt-3′) is different from an 807 bp region in the MATα locus (5′-tatgtctagtatgctggatttaaa…gttgacgaataattatgctgaagt-3′). We used the CRISPR–Cas9 system to replace the 703 bp sequence with the 807 bp sequence in the BY4742 and SY14 strains, to generate BY4742^a and SY14^a strains, respectively.

Mating and sporulation assays. Mating of MATa and MATα strains was performed on YPD plates by micromanipulation. Colonies of the resulting diploid strains were re-streaked on a new YPD plate to obtain a single clone. The diploids were verified by PCR amplifications using the two pairs of primers specific for the different mating-type loci. For sporulation, diploid strains were inoculated into 5 ml of sporulation medium to a density of $OD_{600} = 0.6$ and cultivate the strains at 23 °C. Use hemocytometer to count a total of $\geq$ 200 cells every 24 h, and sporulation efficiency was measured by the ratio of the number of asci containing 3 or 4 spores to that of the unsporulated cells. Spores were dissected, and at least 60 tetrads were dissected to measure spore viabilities.

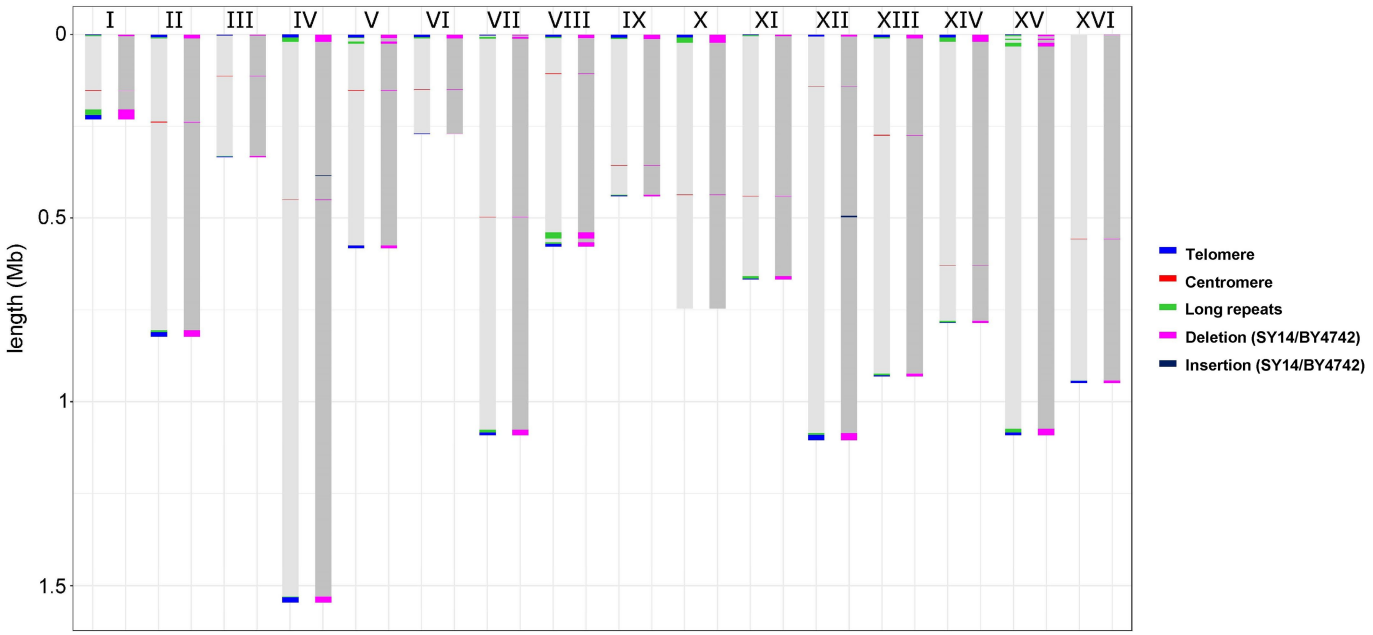
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data Availability. Genome sequencing data and the assembled genome sequences of SY14 and WT have been submitted to NCBI with a project accession number of PRJNA429985. The Hi-C sequencing data of SY6, SY13, SY14 and BY4742 have been submitted to NCBI with a project accession number of PRJNA431161. The RNA-seq data have been submitted to NCBI with a project accession number of PRJNA451522. All data can be viewed in NODE (http://www.biosino.org/node) by pasting the accession (NODEP00371807) into the text search box or through the URL: http://www.biosino.org/node/project/detail/NODEP00371807.

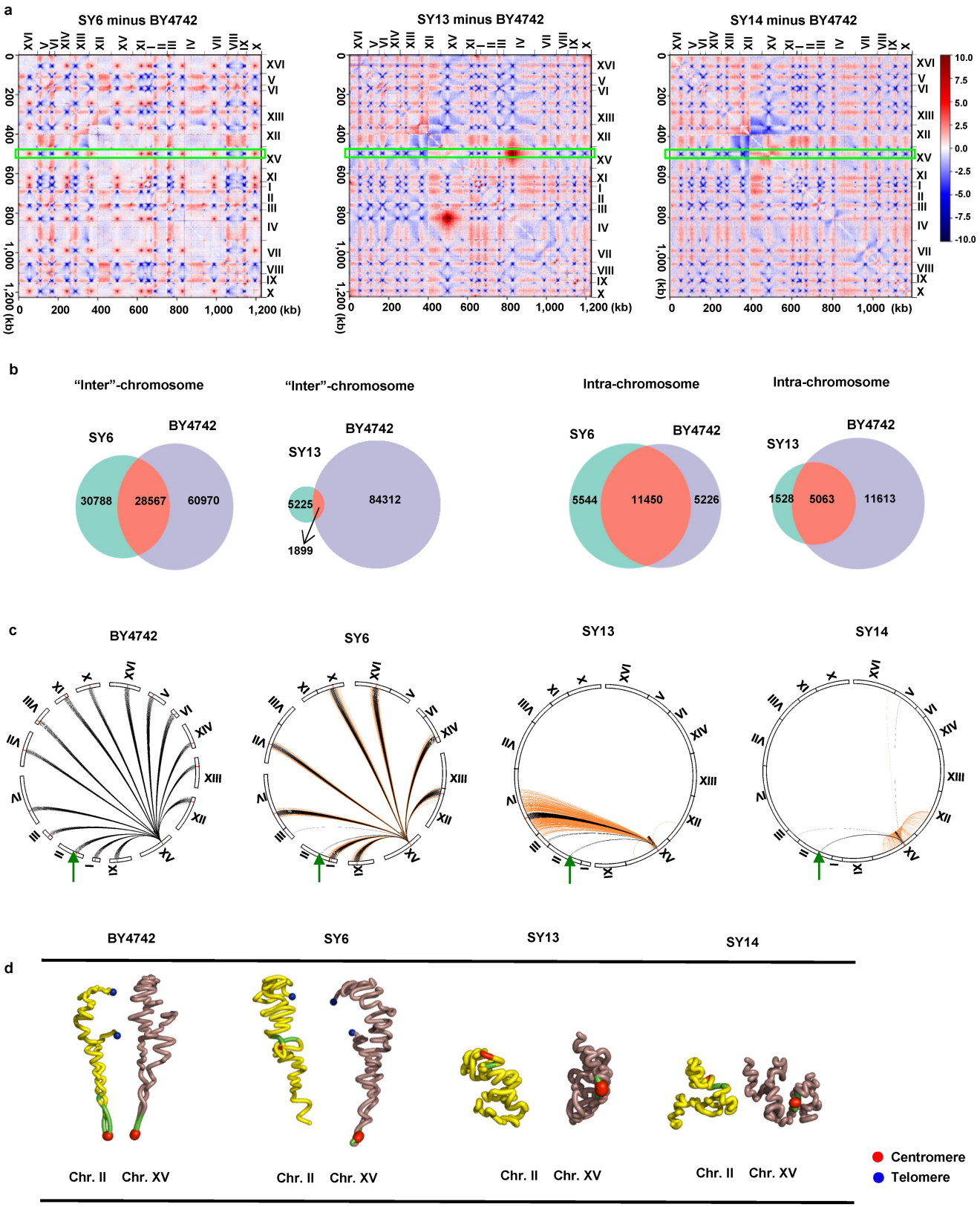
28. Zhou, J., Wu, R., Xue, X. & Qin, Z. CasHRA (Cas9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. *Nucleic Acids Res.* **44**, e124 (2016).
29. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
30. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protocols* **2**, 31–34 (2007).
31. Wu, Z. et al. Rad6-Bre1-mediated H2B ubiquitination regulates telomere replication by promoting telomere-end resection. *Nucleic Acids Res.* **45**, 3308–3322 (2017).
32. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
33. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
34. Frith, M. C. & Kawaguchi, R. Split-alignment of genomes finds orthologies more accurately. *Genome Biol.* **16**, 106 (2015).
35. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
36. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
37. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
38. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
39. Varoquaux, N., Ay, F., Noble, W. S. & Vert, J. P. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* **30**, i26–i33 (2014).
40. Sauria, M. E., Phillips-Cremins, J. E., Corces, V. G. & Taylor, J. HiFive: a tool suite for easy and efficient HiC and 5C data analysis. *Genome Biol.* **16**, 237 (2015).
41. Yardimci, G. et al. Measuring the reproducibility and quality of Hi-C data. Preprint at https://www.biorxiv.org/content/early/2018/02/05/188755 (2017).
42. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
43. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
44. Xie, T., Yang, Q.-Y., Wang, X.-T., McLysaght, A. & Zhang, H.-Y. Spatial colocalization of human ohnolog pairs acts to maintain dosage-balance. *Mol. Biol. Evol.* **33**, 2368–2375 (2016).
45. Gayral, P. et al. Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol. Ecol. Resour.* **11**, 650–661 (2011).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
49. Bennis, S., Chami, F., Chami, N., Bouchikhi, T. & Remmal, A. Surface alteration of *Saccharomyces cerevisiae* induced by thymol and eugenol. *Lett. Appl. Microbiol.* **38**, 454–458 (2004).
50. Manukyan, A., Abraham, L., Dungrawala, H. & Schneider, B. L. Synchronization of yeast. *Methods Mol. Biol.* **761**, 173–200 (2011).



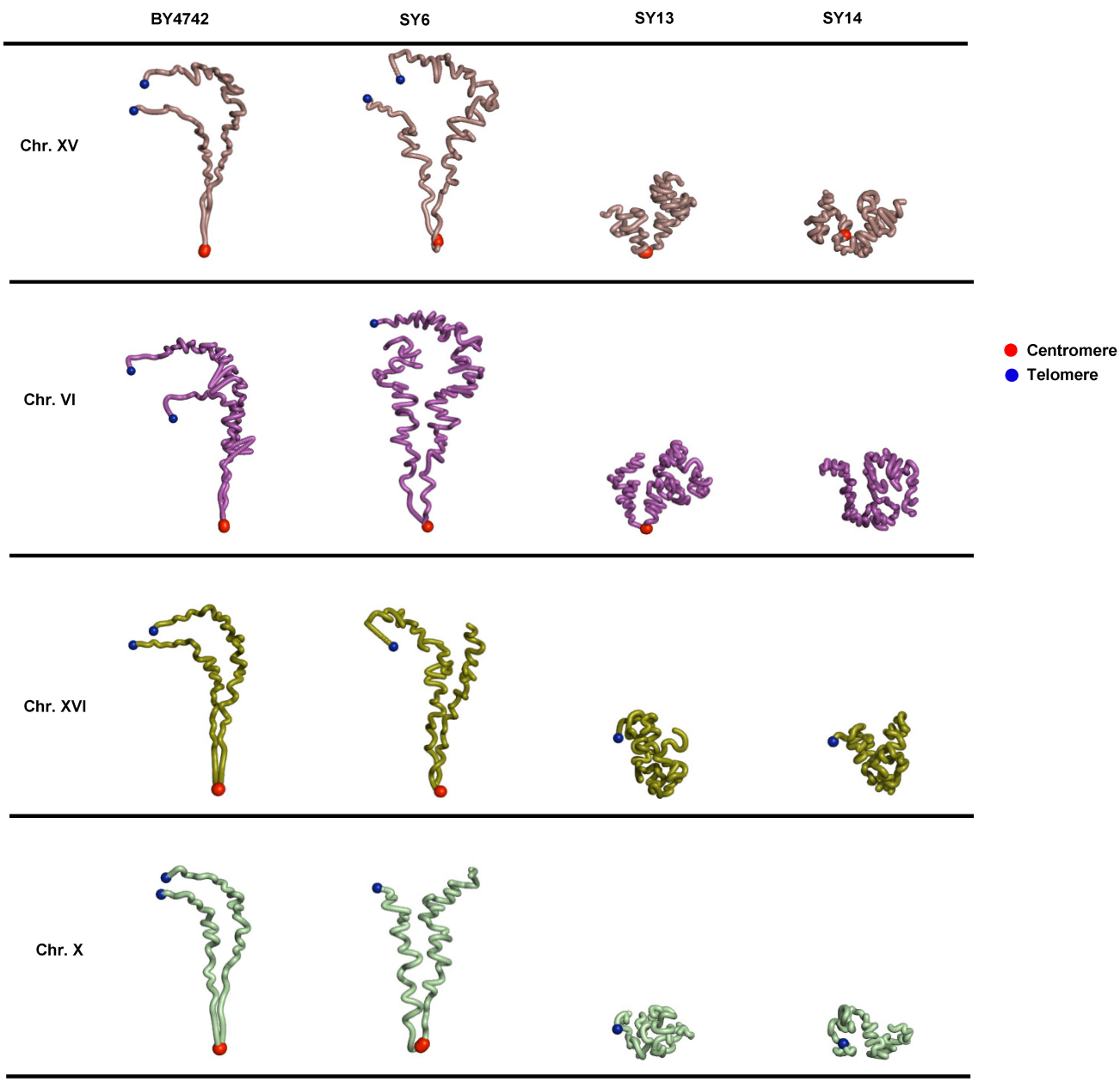
Extended Data Fig. 1 | Theoretical XhoI digestion pattern of chromosome ends. The X, STR, and Y′ elements in each subtelomeric regions are marked with black, grey and white boxes, respectively. The TG_{1-3} telomeric sequences are marked with arrow tips. The XhoI digestion sites in telomere regions are indicated, and the numbers in kb in parenthesis indicate the sizes of DNA fragments recognized by the TG_{1-3} probe in the Southern blot analysis.



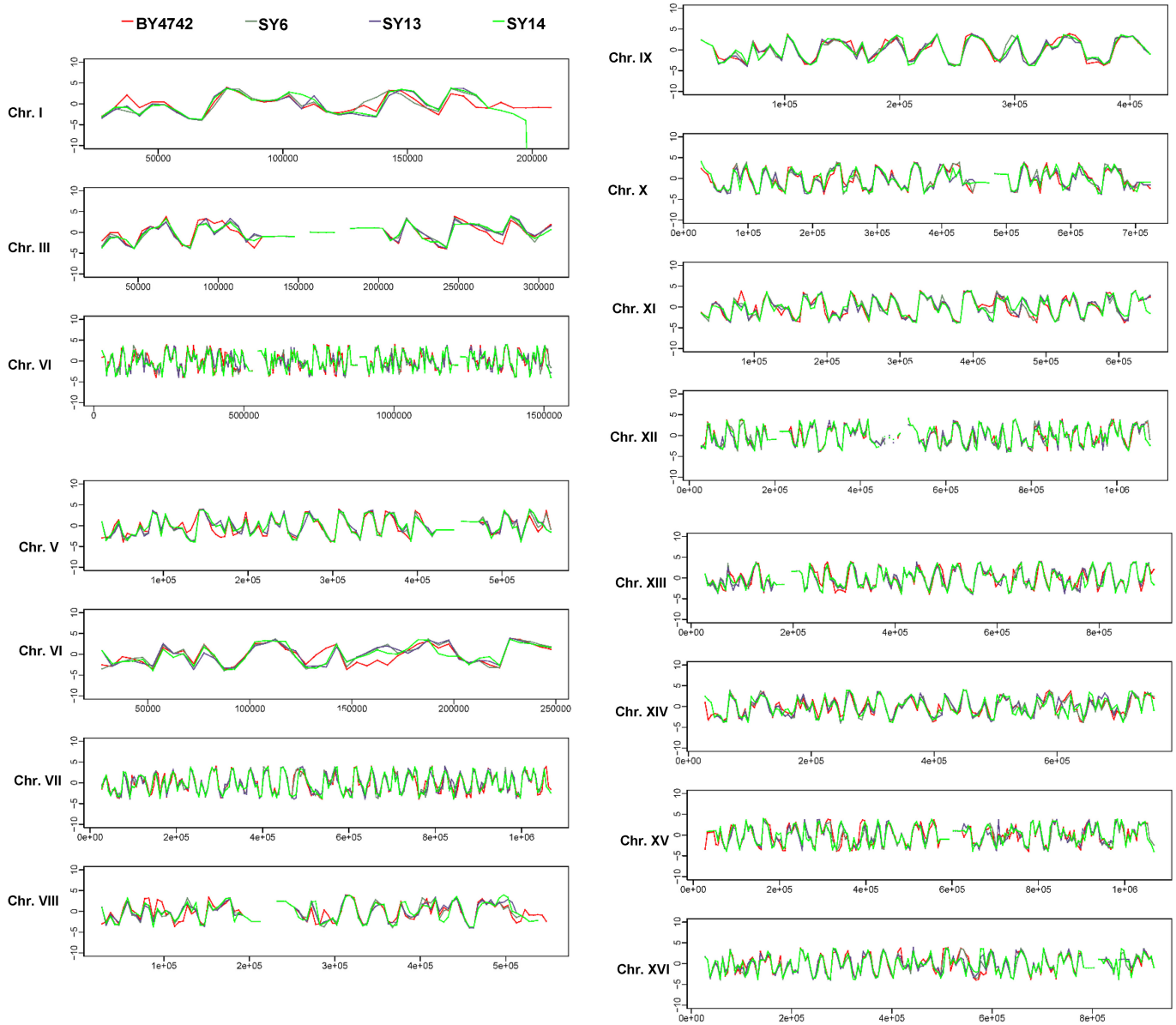
Extended Data Fig. 2 | De novo sequence comparison of BY4742 (light grey) and SY14 (dark grey) genomes. The chromosomes are labelled with Roman numerals of the yeast reference genome. The telomeres (blue), centromeres (red) and telomere-associated repeats (green) that were cut by experimental design are shown in BY4742 chromosomes. Sequence deletions and insertions identified by genomic comparison between BY4742 and SY14 are highlighted in purple and black, respectively, in SY14 chromosomes.



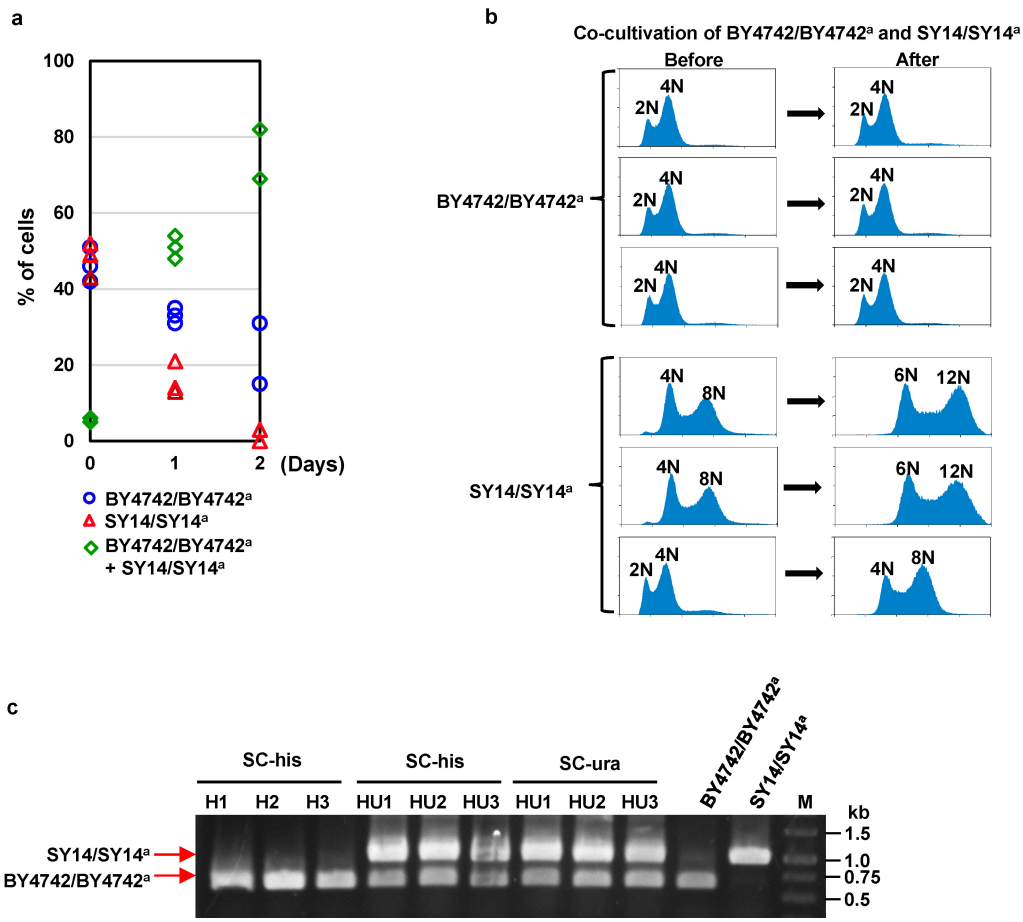
Extended Data Fig. 3 | Comparison of the chromosomal interactions of SY6, SY13 and SY14 cells with those of BY4742 cells. a, *Z*-score difference heatmaps. Bin length, 10 kb; red and blue show increased and decreased chromatin interactions, respectively. Green box highlights the interactions of the chromosome XV centromere with other chromosomes. **b**, Venn diagram of the number of significant ($P < 0.01$, $q < 0.01$) 'inter'- and intra-chromosomal interactions (referring to their locations in the BY4742 genome). **c**, Strong chromosomal interactions of chromosome XV centromere regions in the BY4742, SY6, SY13 and SY14 genomes. The red bars indicate the centromeres and their flanking regions of 50 MboI restriction sites. Each arc throughout the chromosome XV centromere area represents one strong interaction. In SY6, SY13 and SY14, the reserved interactions are marked with black arcs and new interactions are marked with orange arcs. The green arrowheads mark the ten residual interactions near the centromere regions found in all four strains. **d**, 3D structures of chromosomes XV and II in SY6, SY13 and SY14 cells compared to those in BY4742 cells. The locations of the 10 residual interactions on Chr. XV and II are marked green.



Extended Data Fig. 4 | 3D structures of single chromosomes. Chromosome structures in SY6, SY13 and SY14 cells are compared to those in BY4742 cells.



Extended Data Fig. 5 | Directional preference plots of SY6, SY13, and SY14 cells compared to BY4742 cells. Red, BY4742; moss green, SY6; purple, SY13; bright green, SY14. The *y*-axis denotes the *t*-test value between the upstream and downstream interactions of each bin. A positive *t*-value indicates that a bin has more downstream interactions, as described previously[16].



Extended Data Fig. 6 | Growth competition between SY14/SY14ª and BY4742/BY4742ª diploid cells. a, Blue circles represent BY4742/BY4742ª (with *HIS3* marker) cells that could grow only on SC-His plates; pink triangles represent SY14/SY14ª (with *URA3* marker) cells that could grow only on SC-Ura plates; green diamonds represent 'fusion' cells of BY4742/BY4742ª and SY14/SY14ª that could grow on both SC-His and SC-Ura plates. Data from three biological replicates are presented. **b**, FACS analysis of DNA content of BY4742/BY4742ª and SY14/SY14ª diploid cells before and after co-culture. Data are representative of two independent experiments. **c**, PCR verification of genomes from BY4742/BY4742ª and SY14/SY14ª diploid cells. H1–H3: colonies grown only on SC-His plates; HU1–3: colonies grown on both SC-His and SC-Ura plates. The BY4742/BY4742ª and SY14/SY14ª diploid cells before co-cultivation were used as control. Two pairs of primers, specific for genomes of BY4742/BY4742ª and SY14/SY14ª, were used. Data shown are representative images of two independent experiments.

Extended Data Table 1 | Details of the creation of a single chromosome yeast

Strain	Chromosomes to fuse (chr. length in kb)	Newly fused chromosome (+) (chr. length in kb)	Newly deleted chromosome regions (RTel: right arm telomere sequence; LTel: left arm tel sequence) (Cen: centromere sequence; RS: repetitive sequence)	No. of transformants for chromosome fusion	Positive rate of chromosome fusion
SY0	VII (1090), VIII (560)	VII+VIII (1609)	VII: 5545-9584 (RS6), 1076068-1090940 (RTel, RS3); VIII: 524501-541100 (RS1), 1-8217 (LTel), 105447-106013 (Cen) V: 18067-23447 (RS11); XV: 10454-13126 (RS9), 21791-30776 (RS10,14);	100	1/3
SY1	XIII (920), XII (1080+1500)	XIII+XII (1980+1500)	XIII: 917281-924431 (RTel, RS7), 268013-268803 (Cen); XII: 1-14474 (LTel)	26	1/3
SY2	I (230), II (810)	I+II (1006)	I: 203183-230218 (RTel, RS1,2); II: 1-9089 (LTel), 237784-238794 (Cen)	11	2/3
SY3	VI (270), XIV (780)	VI+XIV (1036)	VI: 269616-270161 (RTel), 148460-148774 (Cen); XIV: 1-17790 (LTel, RS4)	100	2/3
SY4	XVI (950), V (580)	XVI+V (1505)	XVI: 941976-948066 (RTel); V: 1-8079 (LTel), 151829-152588 (Cen)	143	4/8
SY5	IX (440), X (750)	IX+X (1160)	IX:436361-439888 (RTel), 355607-356006 (Cen); X:1-21750 (LTel, RS9,10,15)	94	1/4
SY6	III (320), IV (1530)	III+IV (1826)	III:313621-316620 (RTel), 114297-114969 (Cen); IV:1-19188 (LTel, RS12)	49	2/4
SY7	XVI-V (1505), VI-XIV (1036)	XVI-V+VI-XIV (2523)	V:569325-576874 (RTel); VI:1-8380 (LTel); XIV:628734-629219 (Cen)	130	1/5
SY8	XV (1090), XI (670)	XV+XI (1725)	XV:1073966-1091291 (RTel, RS8,7); XI: 439551-440264 (Cen), 1-3182 (LTel)	84	3/3
SY9	VII-VIII (1609), IX-X (1160)	VII-VIII+IX-X (2747)	VIII:552000-562643 (RTel); IX:1-11214 (LTel); X:436229-436425 (Cen)	30	2/5
SY10	I-II (1006), III-IV (1826)	I-II+III-IV (2821)	I:151470-151745 (Cen); II:809896-813184 (RTel); III:1-1500 (LTel)	5	2/4
SY11	XIII-XII (1980+1500), XV-XI (1725)	XIII-XII+XV-XI (3682+1500)	XII:1059256-1078177 (RTel, RS13), 150705-150969 (Cen); XV:1-3709 (LTel)	100	3/5
SY12	I-II-III-IV (2821), VII-VIII-X-X (2747)	I-II-III-IV+VII-VIII-IX-X (5558)	IV:1522800-1531933 (RTel); VII:1-1240 (LTel), 496756-497069 (Cen)	100	2/4
SY13	XVI-V-VI-XIV (2523), XIII-XII-XV-XI (3682+1500)	XVI-V-VI-XIV+XIII-XII-XV-XI (6191+1500)	XIV:779113-784333 (RTel); XVI:555957-556549 (Cen); XIII:1-9319 (LTel)	17	4/4
SY14	XVI-V-VI-XIV-XIII-XII-XV-XI (6191+1500), I-II-III-IV-VII-VIII-IX-X (5558)	XVI-V-VI-XIV-XIII-XII-XV-XI+I-II-III-IV-VII-VIII-IX-X (11737+1500)	IV:449389-449890 (Cen); XI:658549-666816 (RTel, RS5); I:1-3294 (LTel)	26	3/5

The 'Strain' column lists the strain names, and the number in parentheses indicates the size of the native or fused chromosome in kilobase (kb). An orange plus indicates a fusion event; a dash between two chromosomes means that the fusion already occurred. 'Newly deleted chromosome regions' marks the deleted regions in the corresponding chromosomes; Rtel and Ltel in blue indicate the right arm and left arm of the corresponding telomere sequences, respectively; Cen in red indicates the corresponding centromere sequence; and RS represents repetitive sequences deleted in the corresponding chromosomes. The numbers for each region are referred from the *S. cerevisiae* S288C genome (http://www.yeastgenome.org/).

Extended Data Table 2 | Information regarding long repeat sequences near chromosome ends

Types of repeat sequences	Copy number	Location on chromosomes (bp)	
		RSs to be deleted	Retained RSs
RS1	3	I: 203448-219229; VIII: 525437-539926	I:13089-27923
RS2	2	I:219230-229411	VIII: 539927-543610 ; 549638- 556001
RS3	2	VII: 1076381-1083886	II: 805133-812631
RS4	2	XIV: 7429-15942	VI: 5531-14039
RS5	2	XI: 658572-665429	III: 4327-11225
RS6	2	VII: 6223-9584	IX: 430983-434367
RS7	3	XIII: 917474-923540; XV: 1078061-1083736	XVI: 7409-13083
RS8	2	XV: 1073988-1078544	XVI: 12601-17099
RS9	3	XV: 11053-13126; X: 8269-10330	IX: 8286-10347
RS10	3	XV:22397-27006; X: 16639-21229	IX:16656-21250
RS11	2	V: 18751-23447	XIV: 772693-777126
RS12	2	IV: 905-18681	X: 727164-744901
RS13	2	XII: 1059296-1064280	III: 303903-308316
RS14	2	XV: 27007-30776	IX: 21251-25254
RS15	2	X:10331-16638	IX: 10348-16655

This Table lists 15 types of long (>2 kb) repeat sequences near telomeres, which have two or three copies. Only one copy of each long repeat sequence was retained, and the redundant copies were deleted in the SY14 strain. During the generation of the SY14 strain, six long repeats marked in red (that is, VII: 6223–9584 (RS6), VIII: 525437–539926 (RS1), V: 18751–23447 (RS11), XV: 11053–13126 (RS9), XV: 22397–27006 (RS10), and XV: 27007–30776 (RS14)), which are distal to telomeres, were deleted by two rounds of CRISPR–Cas9-mediated PCR targeting. The remaining 13 long repeats were deleted during chromosomal end-to-end fusions.

Extended Data Table 3 | SNPs and indels confirmed by re-sequencing

a

Ref-Chromosome	Ref-Loci	Ref base	Illumina sequencing		Sanger sequencing		Affected Gene	Variation Loci in Gene	Ref Codon	Variation Codon	Variation Type	Mutation in functional domain? (Y/N)	Null gene phenotypes
			WT	SY14	WT	SY14							
II	9136#	G	G	A	G	A	--	--	--	--	--	--	--
III	151645	G	G	A	G	A	--	--	--	--	--	--	--
X	517916	G	G	T	G	T	--	--	--	--	--	--	--
XV	161691	C	C	G	C	G	--	--	--	--	--	--	--
IX	145512	G	G	T	G	T	Nup159	3198	CTC->L	CTA->L	synonymous	--	--
II	447712	A	A	T	A	T	Sif2	4	AGT->S	TGT->C	non-synonymous	N	--
VII	775929	C	C	A	C	A	Skn1	737	GCT->A	GAT->D	non-synonymous	N	--
XV	454309	C	C	A	C	A	Vps5	542	ACA->T	AAA->K	non-synonymous	N	--
X	148553	G	G	A	G	A	Yak1	1838	TCT->S	TTT->F	non-synonymous	Y. Mutation within kinase domain	Null mutant grows slowly, has increased fitness and lifespan, is sensitive to DNA damage.
XI	594128	C	C	A	C	A	Nup133	1304	TCT->S	TAT->Y	non-synonymous	Y. Mutation within WD40/YVTN repeat-like-containing domain	Null mutant has abnormally elongated bud morphology, decreased cell death, abnormal chemical compound accumulation
XIV	225122	A	A	G	A	G	Sqs1	1978	TGG->W	CGG->R	non-synonymous	Y. Mutation within R3H domain	Null mutant has decreased or increased competitive fitness depend on the growth conditions.

b

Ref-Chromosome	Ref-Loci	Ref base	ALTbase	Ref len	ALT len	Illumina sequencing		Sanger sequencing		Affected Gene	Variation Loci in Gene	Mutation in functional domain? (Y/N)	Null gene phenotypes
						WT	SY14	WT	SY14				
III	114269*	TCCCC	TCCC	5	4	0	-1	0	-1	--	--	--	--
IV	450188#	GTATATATATATATATATAT	GTATATATATATATATAT	24	20	0	-4	0	-4	--	--	--	--
VIII	524500*	T	TA	1	2	0	1	0	1	--	--	--	--
XIII	9319*	CTTT	CTTTT	4	5	0	1	0	1	--	--	--	--
XIV	17812*	TC	T	2	1	0	-1	0	-1	--	--	--	--
XVI	941970*	TG	T	2	1	0	-1	0	-1	--	--	--	--
XV	1037517	CAATAATAATAATAATAATAATAATAATA	CAATAATAATAATAATAATAATAATAATAATA	35	38	0	3	0	3	Nud1p	684	N	--

#: errors located within recombination regions during chromosome fusions, but not located within primer binding sites.
***: errors located within primer binding sites.**

a, SNPs. **b**, Indels.

Extended Data Table 4 | Differentially expressed genes in SY14 compared to BY4742 cells

geneID	GeneSymbol	Expression						log_2FC	Pvalue	FDR	Notes
		SY14-1	SY14-2	SY14-3	WT-1	WT-2	WT-3				
Telomere adjacent genes (<10 kb distance)											
855848	ERR2	0.04	0	0.04	2.08	1.52	2.05	-7.55	7.67E-13	6.78E-11	Chr XVI-L.
855849	HSP32	0	0.1	0	2.86	2.6	2.57	-6.83	2.83E-09	2.03E-07	Chr XVI-L.
855850	FEX2	1.11	4.51	2.73	21.54	21.83	23.7	-3.03	1.64E-12	1.38E-10	Fourth gene near XVI-L.
855852	YPL277C	1.97	2.82	2.09	7.11	5.57	8.61	-1.66	0.000071	0.004003	Sixth gene near XVI-L.
853625	MPH3	3.79	7.58	7.24	19.12	18.21	19.61	-1.65	2.78E-05	0.001666	Second gene near X-R.
854634	VTH1	25.96	23.78	29.85	8.71	8.08	11.03	1.47	0.000126	0.006776	Second gene near IX-L, first gene was deleted in SY14.
851230	SEO1	6.79	7.37	7.27	2.72	2.02	2.69	1.49	0.000254	0.013032	Third gene near I-L, the first two genes were deleted in SY14.
854002	YOL162W	25.86	21.03	21.94	10.17	6.31	5.28	1.64	9.25E-05	0.005119	Fourth gene near XV-L.
854001	YOL163W	42.33	35.02	31.7	14.94	7.6	8.11	1.82	1.53E-05	0.000936	Third gene near XV-L.
850486	THI5	4.61	5.13	4.56	1.51	0.14	1.93	1.93	3.53E-05	0.002072	Fifth gene near VI-L, first gene was deleted in SY14.
853207	MAL11	59.06	56.38	56.36	8.7	9.65	6.49	2.75	1.04E-11	8.32E-10	Fourth gene near VII-R,first three genes were deleted in SY14.
850618	YFR057W	22.58	20.31	18.16	3.13	1.46	0.96	3.40	1.64E-10	1.2E-08	First gene near VI-R.
Stress response											
852364	HSP26	60.43	56.15	60.09	24.96	18.59	20.85	1.44	0.000263	0.013364	Small heat shock protein (sHSP) with chaperone activity.
854744	RNR3	15.67	14.26	15.89	6.2	4.95	5.09	1.46	0.000184	0.009714	Minor isoform of large subunit of ribonucleotide-diphosphate reductase;regulated by DNA replication and DNA damage checkpoint pathways. Induced by DNA damage and replication stress.
853326	NCA3	85.44	90.16	83.51	32.42	31	28.38	1.47	0.000148	0.007853	Protein involved in mitochondrion organization; Induced by the SLN1-SKN7 osmotic stress signaling pathway.
850331	HBN1	30.17	22.64	28.11	10.71	6.16	8.97	1.63	9.49E-05	0.005147	Similar to bacterial nitroreductases; protein abundance increases in response to DNA replication stress.
855932	OYE3	8.01	5.63	9.95	1.71	2.57	2.83	1.69	7.08E-05	0.004003	Conserved NADPH oxidoreductase containing flavin mononucleotide (FMN),has potential roles in oxidative stress response and programmed cell death.
854944	HUG1	479.95	1084.46	860.2	381.3	251.63	167.35	1.82	0.00048	0.023808	Ribonucleotide reductase inhibitor; transcription is induced by genotoxic stress and by activation of the Rad53p pathway; protein abundance increases in response to DNA replication stress.
850532	HSP12	211.2	223.37	183.79	62.8	54.47	46.85	1.98	1.56E-06	0.000102	Plasma membrane protein involved in maintaining membrane organization; involved in maintaining organization during stress conditions; induced by heat shock, oxidative stress, osmostress,; protein abundance increased in response to DNA replication stress.
855132	YKU80	12.19	12.28	12.7	0	0.06	0.1	7.65	1.06E-33	1.97E-31	Subunit of telomeric Ku complex (Yku70p-Yku80p); involved in telomere length maintenance, structure and telomere position effect; relocates to sites of double-strand cleavage to promote nonhomologous end joining during DSB repair.
Others											
9164931	RDN37-1	82.96	93.16	86.11	13.39	13.71	16.74	2.54	1.53E-10	1.13E-08	35S ribosomal RNA (35S rRNA) transcript.
854565	FIT3	489.5	230.81	526.59	265.15	96.71	154.28	1.25	0.00102	0.049686	Mannoprotein that is incorporated into the cell wall.
852298	YBR012W-A	20.41	15.07	15.39	5.28	2.1	5.84	1.91	2.22E-06	0.000143	Retrotransposon TYA Gag gene.
854004	YOL160W	15.64	15.33	26.41	1.1	3.52	1.28	3.23	9.57E-07	6.39E-05	Putative protein of unknown function; conserved across *S. cerevisiae* strains.
851850	YDR261W-A	0.24	0	0	1.2	1.94	0	-3.57	3.33E-06	0.000213	Transposable_element_gene.
850849	TAR1	12.98	0	0.08	25.05	66.46	0	-2.88	3.81E-08	2.6E-06	Transcript Antisense to Ribosomal RNA. Protein potentially involved in regulation of respiratory metabolism.
854860	YIR042C	10.6	10.01	8.16	53.09	44.8	51.94	-2.40	5.57E-09	3.95E-07	Second gene near IX-R, first gene deleted in SY14.
853869	MET14	40.28	38.8	41.33	127.39	96.51	128.82	-1.56	5.97E-05	0.00343	First gene near centromere.

Sample size $n = 3$. Exact negative binomial two-sided test was used to generate *P* values. Benjamini and Hochberg's algorithm was used to control the FDR.

LETTER

https://doi.org/10.1038/s41586-018-0336-3

Ferroelectric switching of a two-dimensional metal

Zaiyao Fei[1,5], Wenjin Zhao[1,5], Tauno A. Palomaki[1,5], Bosong Sun[1], Moira K. Miller[1], Zhiying Zhao[2,3], Jiaqiang Yan[2], Xiaodong Xu[1,4] & David H. Cobden[1]*

A ferroelectric is a material with a polar structure whose polarity can be reversed (switched) by applying an electric field[1,2]. In metals, itinerant electrons screen electrostatic forces between ions, which explains in part why polar metals are very rare[3–7]. Screening also excludes external electric fields, apparently ruling out the possibility of ferroelectric switching. However, in principle, a thin enough polar metal could be sufficiently penetrated by an electric field to have its polarity switched. Here we show that the topological semimetal WTe_2 provides an embodiment of this principle. Although monolayer WTe_2 is centro-symmetric and thus non-polar, the stacked bulk structure is polar. We find that two- or three-layer WTe_2 exhibits spontaneous out-of-plane electric polarization that can be switched using gate electrodes. We directly detect and quantify the polarization using graphene as an electric-field sensor[8]. Moreover, the polarization states can be differentiated by conductivity and the carrier density can be varied to modify the properties. The temperature at which polarization vanishes is above 350 kelvin, and even when WTe_2 is sandwiched between graphene layers it retains its switching capability at room temperature, demonstrating a robustness suitable for applications in combination with other two-dimensional materials[9–12].

A polar material contains an axis (referred to as the polar axis) along which the two opposite directions are distinguishable. This property is necessary for the existence of a spontaneous electric polarization. Of the 32 three-dimensional crystal classes, the ten that have a polar axis are known as the pyroelectrics, because heating them changes any electric polarization along this axis to produce a voltage. When Anderson and Blount introduced[3] the term 'ferroelectric metal' in 1965, they were referring to the possibility of polar structure appearing in certain metallic crystals upon cooling. However, they assumed that, even if such polar metals existed, the polarity would not be switchable. Definite cases of metals with polar structure have been identified only very recently[4–7].

Several ferroelectric insulators have been found to maintain ferroelectric characteristics in ultrathin films[13–17]. However, when materials with a layered structure are thinned towards the monolayer limit their properties often change qualitatively. This is illustrated by, for example: graphene, which becomes a two-dimensional Dirac metal[11,12]; MoS_2, which changes from an indirect- to a direct-gap semiconductor[10]; and CrI_3, which varies between being antiferromagnetic and ferromagnetic[9]. Another example is the topological semimetal WTe_2[18], which becomes either a two-dimensional topological insulator[19–22] or a superconductor at low temperatures in the monolayer limit, depending on the level of electrostatic doping. Here we focus on another aspect of WTe_2: the fact that it is a polar metal. Its three-dimensional (1T′) structure has a polar space group[7], $Pnm2_1$, and it remains metallic down to a thickness of three layers when undoped[23] and a monolayer when electrostatically doped[20]. We show here that as WTe_2 approaches this limit the polarity can be switched, making it effectively ferroelectric even when it is metallic in the plane.

The 1T′ structure (Fig. 1a) contains *b*–*c* mirror (M) and *a*–*c* glide (G) planes, so the polar axis, which must be parallel to both of them, is the *c* axis, perpendicular to the layers[7,18]. We apply an electric field along this axis using the device geometry indicated in Fig. 1b. An electrically contacted thin WTe_2 flake is sandwiched between two hexagonal boron nitride (h-BN) dielectric sheets, with thicknesses of d_t (top) and d_b (bottom). Above and below are gate electrodes, usually of few-layer graphene, to which voltages V_t and V_b are applied relative to the grounded WTe_2 (see Methods, Extended Data Fig. 1 and Extended Data Table 1 for device fabrication and characterization).

We define the applied electric field passing upwards through the layer, which will couple to out-of-plane polarization, as $E_\perp = (-V_t/d_t + V_b/d_b)/2$. When $E_\perp$ is swept up and down, in the conductance of trilayer (Fig. 1c) and bilayer (Fig. 1d) devices we observe bistability near $E_\perp = 0$, characteristic of ferroelectric switching, at all temperatures T from 4 K to above room temperature. No bistability is seen in monolayer WTe_2 (Fig. 1e), consistent with its structure having a centre of symmetry (Fig. 1e inset, red circles) and hence being non-polar; this symmetry also rules out instabilities involving charge injection into the h-BN. Nor is bistability seen in thicker crystals, including when one is used as a gate electrode (Extended Data Fig. 2). This, and the larger field required to switch the trilayer device than the bilayer device, can be explained by screening of $E_\perp$ on a length scale of nanometres.

We saw similar bistability in all bilayer devices (Extended Data Fig. 3). To prove that it is associated with out-of-plane electric polarization, we made devices in which the top gate is replaced by monolayer graphene, the conductivity of which is sensitive to the precise electric field E_t in the upper h-BN. In Fig. 2 we present measurements at a series of temperatures on such a bilayer WTe_2 device (B2) with four gold contacts to the top graphene (Fig. 2a; Extended Data Fig. 4). If the WTe_2 acts as a conducting sheet then it will screen out any electric field due to a voltage applied to the bottom gate. Indeed, Fig. 2b demonstrates that the conductance G_{gr} of the graphene depends only very weakly on V_b, except in a certain interval where it jumps between two states. The conductance of the WTe_2 is bistable in precisely the same interval (Extended Data Fig. 4). The two states must correspond to different values of E_t that can occur for exactly the same set of applied bias voltages. This implies the existence of two different vertical distributions of charge in the bilayer WTe_2. We deduce that sweeping the bottom gate changes $E_\perp$ (here $E_\perp = V_b/(2d_b)$ because $V_t = 0$), which at the ends of the hysteresis loop flips the polarization state (henceforth denoted by P↑ or P↓), changing E_t by an amount δE_t, and so changing G_{gr}.

We infer δE_t by applying a bias V_W to the WTe_2 and measuring the change $\delta V = d_t \delta E_t$ that is required to produce the same change in G_{gr} (Fig. 2c). For the simplified case of $d_t = d_b$ and all voltages at zero, the electrostatic potential profile is inverted between P↑ (red) and P↓ (green), as sketched in Fig. 2d, and the areal polarization density is $P \approx \varepsilon_0 \delta V$ (Methods), where ε_0 is the vacuum permittivity. At 20 K, this gives $P \approx 1 \times 10^4$ e per cm, which is equivalent to transferring about 2×10^{11} electrons per cm^2 between the two layers, a distance of about 0.7 nm. This is three orders of magnitude lower than the volume polarization density of around 0.2 C m^{-2} $\approx 10^{14}$ electrons per cm^2 in the classic ferroelectric[1] $BaTiO_3$. Combined with the micrometre-scale

[1]Department of Physics, University of Washington, Seattle, WA, USA. [2]Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [3]Department of Physics and Astronomy, University of Tennessee, Knoxville, TN, USA. [4]Department of Materials Science and Engineering, University of Washington, Seattle, WA, USA. [5]These authors contributed equally: Zaiyao Fei, Wenjin Zhao, Tauno A. Palomaki. *e-mail: cobden@uw.edu

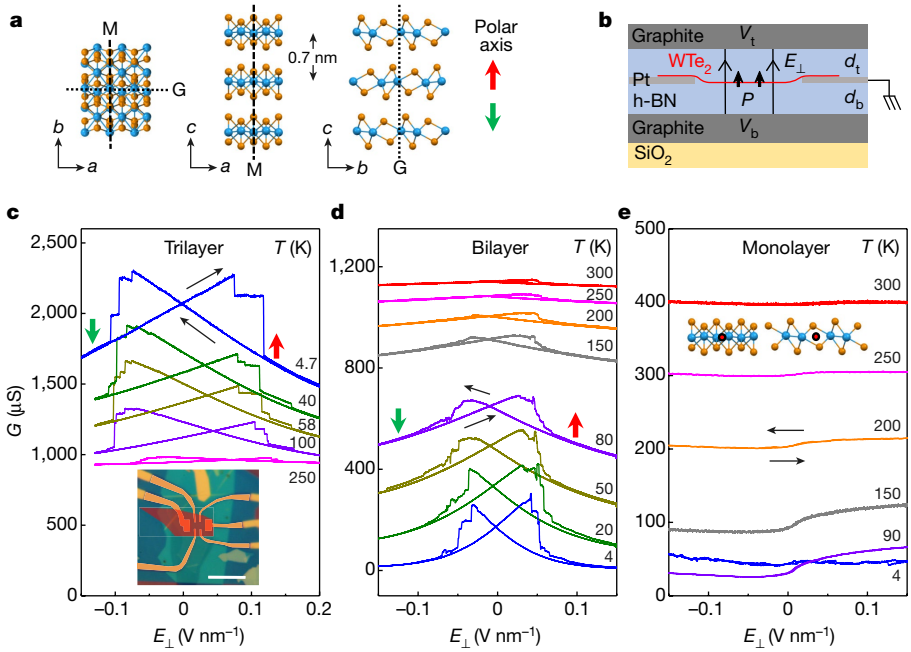


Fig. 1 | Evidence for ferroelectric switching in WTe_2. a, Structure of three-dimensional 1T′ WTe_2, showing the mirror plane (M; dashed), glide plane (G; dotted) and polar *c* axis (red arrow, up; green arrow, down). W atoms are blue; Te atoms are orange. **b**, Schematic cross-section of the device geometry used to apply an electric field $E_\perp$ normal to an atomically thin WTe_2 flake. **c**, **d**, Conductance *G* of undoped trilayer device T1 (**c**) and bilayer device B1 (**d**) as $E_\perp$ is swept up and down (black arrows), setting $V_t/d_t = -V_b/d_b$ to avoid net doping. The plots show bistability associated with electric polarization up (red arrow) or down (green arrow), at temperatures from 4 K to 300 K (as labelled). Here the conductance is the reciprocal of the four-terminal resistance. The undoped trilayer has a metallic temperature dependence, the bilayer an insulating one. Inset to **c**, optical image of a representative double-gated device. The WTe_2 flake has been artificially coloured red. Scale bar, 10 μm. **e**, Similar measurements on a monolayer WTe_2 device (M1), showing no bistability. At 4 K, conduction is in the quantum spin Hall regime. Insets, location (red circle) of a centre of symmetry in the monolayer, viewed along the *b* (left) and *a* (right) axes.

device size, such a small polarization makes it very hard to detect the ferroelectricity using standard displacement current measurements.

In Fig. 2e we plot δV as a function of *T*. Between about 60 K and 300 K it decreases roughly linearly with *T*, extrapolating to zero at roughly 450 K. However, above about 340 K the signal becomes unstable and we can no longer identify a hysteresis loop, suggesting that a transition to a non-polar state occurs in this temperature range.

We also made a simpler device with no top h-BN and monolayer graphene directly encapsulating the bilayer WTe_2. It exhibited highly reproducible hysteresis in the conductance, visible up to 300 K (Extended Data Fig. 5). This result demonstrates that the ferroelectric switching is robust enough for potential applications at room temperature that use it in combination with other two-dimensional materials.

We also investigated the effect of gate-induced charge doping, defined by $n_e = \varepsilon_{hBN}\varepsilon_0(V_t/d_t + V_b/d_b)/e$, where *e* is the electron charge and $\varepsilon_{h\text{-}BN}$ is the relative permittivity of h-BN. If the material were a simple metal, n_e would be the areal density of added electrons. In Fig. 3a, b we plot the conductance *G* at 7 K for device B1 (the same bilayer as in Fig. 1d), as a joint function of V_t and V_b, measured with V_t stepped and V_b swept up or down. Each sweep was started in the same fully polarized state. The black dashed lines in Fig. 3a, b denote $E_\perp = 0$ and the white dashed lines denote $n_e = 0$. The two plots differ only in the central hysteretic region, as is made clearer by plotting the difference between them (Fig. 3c). Similar behaviour is seen at higher temperature (Fig. 3d, at 200 K). At $E_\perp = 0$, *G* is a similar function of n_e for both P↑ and P↓ (Fig. 3e), with a temperature dependence that is insulating near $n_e = 0$ and metallic for $n_e > n_c$, with critical density $n_c \approx 2 \times 10^{12}$ cm^{-2}, as reported previously[20]. In Fig. 3f we show traces obtained by sweeping $E_\perp$ repeatedly up and down for selected values of n_e at 7 K. In each case the single conductance level at large $E_\perp$ evolves smoothly and reproducibly into one of the two stable levels as $E_\perp$ is reduced to zero, implying that the state remains uniformly polarized, without domain structure, at $E_\perp = 0$. For small or negative n_e, the effect of $E_\perp$ is large and of opposite sign for P↑ and P↓, producing butterfly-shaped hysteresis loops. For n_e well above n_c, $E_\perp$ has less effect on the conductance and the hysteresis is smaller, but still present. Hence, the doped bilayer device, like the trilayer device, is simultaneously ferroelectric and metallic.

At low temperatures (Fig. 3c) we observe an increase in the width of the hysteresis loop for increasingly negative n_e, whereas at 200 K (Fig. 3d) it is almost independent of n_e. When the conductance jumps there is some stochastic variation in the positions and substructure of the jumps, which is indicative of domain dynamics. If the surrounding gates were not present to screen the depolarization field, domains would inevitably form to limit the electrostatic energy[24], as observed in other ultrathin ferroelectrics[13,16,25,26]. In our devices, defects such as rips, bubbles and folds could nucleate domains or pin domain walls. In addition, $E_\perp$ is not completely uniform, because above and near the platinum contacts it is reduced by screening. Indeed, the pattern of switching depends on the choice of measurement contacts within a given device (Extended Data Fig. 6). We also observe that in some bilayer devices, such as B2 (Fig. 2), the switching field is not symmetric about $E_\perp = 0$. A possible explanation for this is that sometimes, despite all precautions, during device fabrication one side of the WTe_2 flake was exposed to mild oxidation, producing asymmetric trapped charge.

The fact that the conductance is sensitive to the polarization is consistent with the expectation that the polarization redistributes charge between the layers, which are inequivalent when $E_\perp$ is non-zero. Although the specific mechanisms for the sensitivity to n_e, $E_\perp$ and *P* are still under investigation, we remark on the following. First, the monolayer conductance at 4 K in Fig. 1e, which we know is due to edge conduction because this is the established quantum spin Hall regime, is almost independent of $E_\perp$. Second, in bilayers at large positive or negative n_e, the reversal of *P* has a similar effect on the conductance to that of changing $E_\perp$, changing it by approximately 0.15 V nm^{-1} at 7 K (indicated by the dotted horizontal line in Fig. 3f). This change in $E_\perp$ corresponds to a change in the electrostatic potential difference between the two WTe_2 layers by about 100 mV. This is of the same order

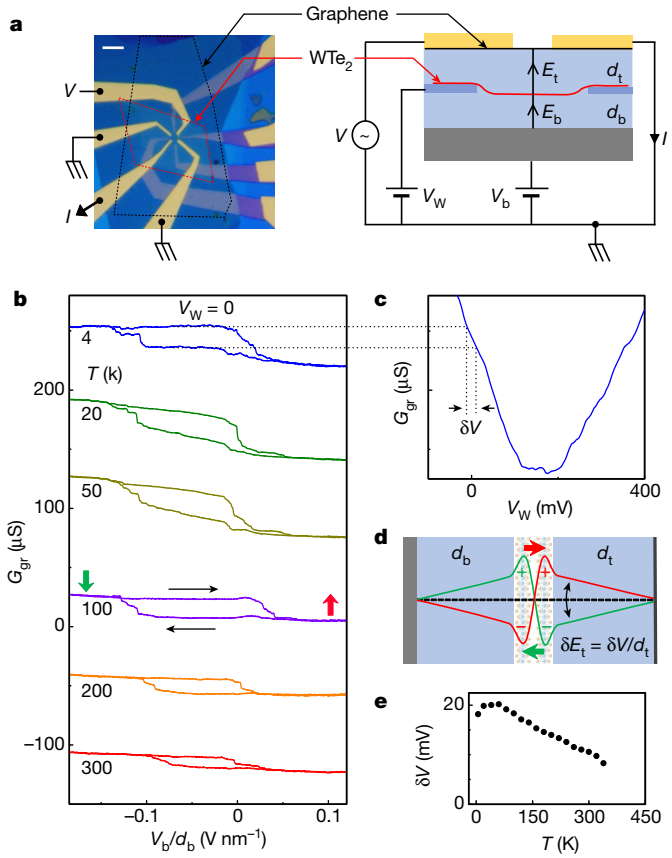


Fig. 2 | Detecting the out-of-plane polarization. a, Micrograph (left) and schematic cross-section (right) of a bilayer WTe_2 device (B2) with multiply contacted graphene in place of the top gate, indicating separately the electric fields in the h-BN above (E_t) and below (E_b) the WTe_2. Scale bar, 5 μm. **b**, The graphene conductance G_{gr} is measured when a bias V_b is applied to the bottom gate with the intervening WTe_2 grounded, at a series of temperatures (as labelled). The two conductance states seen for the two sweep directions (black arrows) are associated with different out-of-plane polarization states of the WTe_2 (red and green arrows as in Fig. 1). **c**, The behaviour of G_{gr} (the y axis is the same as in **b**) when a voltage V_W is applied directly to the WTe_2 provides a mapping to the difference $\delta E_t = \delta V/d_t$ in E_t between the two states. **d**, Sketch indicating how the reversal of the polarization changes the electrostatic potential (from red to green) and E_t (see text). **e**, Temperature dependence of δV, which is proportional to the polarization.

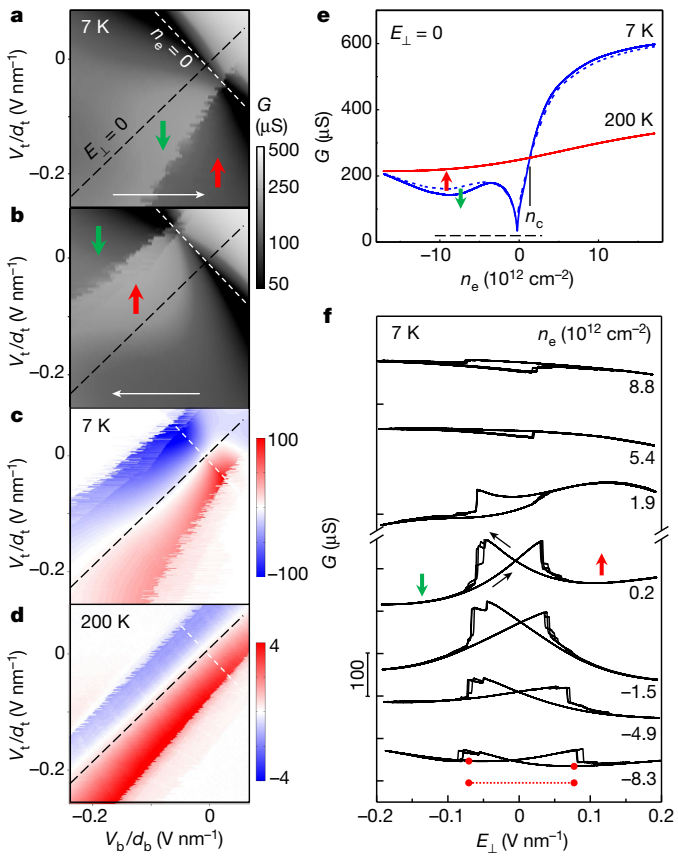


Fig. 3 | Gate tuning of the ferroelectric behaviour. a, **b**, Conductance G of bilayer device B1 at 7 K as a function of both gate voltages, for the two sweep directions of V_b as indicated by the white arrows. **c**, Difference between **a** and **b** at 7 K, which is non-zero in the hysteretic regime. **d**, Same measurement as in **c**, but at 200 K. In **a**–**d**, black and white dashed lines indicate contours of zero perpendicular field $E_\perp$ and zero charge density n_e, respectively. **e**, Variation in G with n_e at $E_\perp = 0$ for both polarization states (up, dashed; down, solid) at two temperatures (as labelled). The dashed bar near the bottom indicates the range of n_e in **a**–**d**, and n_c is the critical density (see text). **f**, Sweeps of $E_\perp$ for fixed n_e (as labelled) at 7 K. The dotted bar near the bottom indicates the magnitude of the approximate shift in $E_\perp$ of the conductance minimum between the opposite polarization states.

as the estimated change in the potential difference associated with the polarization reverse, $2\delta V \approx 40$ mV, suggesting that the potential imbalance between the layers governs the sensitivity of the conductance to both $E_\perp$ and P. It is also roughly the same as the width of the hysteresis loop; that is, the polarization flips roughly when the applied potential difference exceeds the potential due to the spontaneous polarization. This is another indicator that electron transfer between the layers may be involved. Third, the very sharp minimum seen in G close to $n_e = 0$ in bilayers (Fig. 3e) presumably marks the compensation point at which electron and hole densities are exactly equal, suggesting that electron–hole correlation may be important. Taken together, the above observations raise the possibility that electron–hole correlation effects, rather than a lattice instability[27], drive the spontaneous polarization in WTe_2. If this is the case, then the polarization could principally involve a relative motion of the electron cloud relative to the ion cores, rather than a lattice distortion, in which case the switching would be intrinsically very fast.

Ferroelectricity adds another ingredient to the intriguing combination of quantum spin Hall edges, correlation effects and superconductivity already seen in atomically thin WTe_2. Although the quantum spin Hall behaviour and superconductivity are restricted to the centro-symmetric monolayer and ferroelectricity occurs only for two or more layers, it is possible that these diverse phenomena are connected in ways that may also be relevant to understanding the properties that emerge in the three-dimensional limit, including extreme and anisotropic magnetoresistance[28,29], a polar axis and Weyl points[18,30].

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0336-3

Received: 14 February 2018; Accepted: 25 May 2018;
Published online 23 July 2018.

1. Dawber, M., Rabe, K. M. & Scott, J. F. Physics of thin-film ferroelectric oxides. *Rev. Mod. Phys.* **77**, 1083–1130 (2005).
2. Scott, J. F. Applications of modern ferroelectrics. *Science* **315**, 954–959 (2007).
3. Anderson, P. W. & Blount, E. I. Symmetry considerations on martensitic transformations: 'ferroelectric' metals? *Phys. Rev. Lett.* **14**, 217–219 (1965).
4. Shi, Y. et al. A ferroelectric-like structural transition in a metal. *Nat. Mater.* **12**, 1024–1027 (2013).
5. Benedek, N. A. & Birol, T. 'Ferroelectric' metals reexamined: fundamental mechanisms and design considerations for new materials. *J. Mater. Chem. C* **4**, 4000–4015 (2016).
6. Kim, T. H. et al. Polar metals by geometric design. *Nature* **533**, 68–72 (2016).

7. Sakai, H. et al. Critical enhancement of thermopower in a chemically tuned polar semimetal $MoTe_2$. *Sci. Adv.* **2**, e1601378 (2016).
8. Rajapitamahuni, A., Hoffman, J., Ahn, C. H. & Hong, X. Examining graphene field effect sensors for ferroelectric thin film studies. *Nano Lett.* **13**, 4374–4379 (2013).
9. Huang, B. et al. Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit. *Nature* **546**, 270–273 (2017).
10. Mak, K. F., Lee, C., Hone, J., Shan, J. & Heinz, T. F. Atomically thin MoS_2: a new direct-gap semiconductor. *Phys. Rev. Lett.* **105**, 136805 (2010).
11. Novoselov, K. S. et al. Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
12. Zhang, Y., Tan, Y. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
13. Bune, A. V. et al. Two-dimensional ferroelectric films. *Nature* **391**, 874–877 (1998).
14. Martin, L. W. & Rappe, A. M. Thin-film ferroelectric materials and their applications. *Nat. Rev. Mater.* **2**, 16087 (2017).
15. Shirodkar, S. N. & Waghmare, U. V. Emergence of ferroelectricity at a metal-semiconductor transition in a 1T monolayer of MoS_2. *Phys. Rev. Lett.* **112**, 157601 (2014).
16. Liu, F. et al. Room-temperature ferroelectricity in $CuInP_2S_6$ ultrathin flakes. *Nat. Commun.* **7**, 12357 (2016).
17. Fei, R. X., Kang, W. & Yang, L. Ferroelectricity and phase transitions in monolayer group-IV monochalcogenides. *Phys. Rev. Lett.* **117**, 097601 (2016).
18. Soluyanov, A. A. et al. Type-II Weyl semimetals. *Nature* **527**, 495–498 (2015).
19. Qian, X., Liu, J., Fu, L. & Li, J. Quantum spin Hall effect in two-dimensional transition metal dichalcogenides. *Science* **346**, 1344–1347 (2014).
20. Fei, Z. et al. Edge conduction in monolayer WTe_2. *Nat. Phys.* **13**, 677–682 (2017).
21. Tang, S. et al. Quantum spin Hall state in monolayer 1T′-WTe_2. *Nat. Phys.* **13**, 683–687 (2017).
22. Wu, S. et al. Observation of the quantum spin Hall effect up to 100 kelvin in a monolayer crystal. *Science* **359**, 76–79 (2018).
23. Fatemi, V. et al. Magnetoresistance and quantum oscillations of an electrostatically tuned semimetal-to-metal transition in ultrathin WTe_2. *Phys. Rev. B* **95**, 041410(R) (2017).
24. Strukov, B. A. & Levanyuk, A. P. *Ferroelectric Phenomena in Crystals: Physical Foundations* 193–224 (Springer, Berlin, 1998).
25. Catalan, G., Seidel, J., Ramesh, R. & Scott, J. F. Domain wall nanoelectronics. *Rev. Mod. Phys.* **84**, 119–156 (2012).
26. Chang, K. et al. Discovery of robust in-plane ferroelectricity in atomic-thick SnTe. *Science* **353**, 274–278 (2016).
27. Cochran, W. Crystal stability and the theory of ferroelectricity. *Adv. Phys.* **9**, 387–423 (1960).
28. Ali, M. N. et al. Large, non-saturating magnetoresistance in WTe_2. *Nature* **514**, 205–208 (2014).
29. Zhao, Y. F. et al. Anisotropic magnetotransport and exotic longitudinal linear magnetoresistance in WTe_2 crystals. *Phys. Rev. B* **92**, 206803(R) (2015).
30. Wu, Y. et al. Observation of Fermi arcs in the type-II Weyl semimetal candidate WTe_2. *Phys. Rev. B* **94**, 121113 (2016).

Acknowledgements We thank J. Folk, E. Sajadi, A. Levanyuk, T. Birol and A. Andreev for substantial insights. D.H.C. and X.X. were supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under awards DE-SC0002197 and DE-SC0018171, respectively. Synthesis efforts at ORNL were also supported by the same division of the Department of Energy. Z.Z. was partially supported by the CEM, and NSF MRSEC, under grant DMR-1420451. T.A.P. was supported by AFOSR FA9550-14-1-0277. Z.F., W.Z. and B.S. were supported by the above awards and also by NSF EFRI 2DARE 1433496 and NSF MRSEC 1719797.

Reviewer information *Nature* thanks L. Bartels, T. Birol and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.H.C. conceived the experiments; Z.Z. and J.Y. grew the crystals; W.Z., T.A.P., Z.F. and M.K.M. fabricated the devices; Z.F., W.Z., T.A.P. and B.S. performed the measurements; D.H.C., X.X., Z.F., W.Z. and T.A.P. analysed the results; and D.H.C., Z.F., T.A.P. and X.X. wrote the paper with comments from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0336-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.H.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Preparation and characterization of WTe_2 devices. We measured devices with four different layouts: (1) WTe_2 with graphite gates above and below (M1, B1, B4, T1); (2) bilayer WTe_2 with monolayer graphene as a top gate (B2); (3) a bilayer WTe_2/graphene heterostructure (B3); and (4) a monolayer graphene device gated by few-layer WTe_2 (F1). In the following, we describe fabrication of the first type; the others are similar.

First, graphite and h-BN crystals were mechanically exfoliated under ambient conditions onto substrates consisting of 285-nm thermal SiO_2 on highly p-doped silicon. Graphite flakes 2–6 nm thick were chosen for the top and bottom gates and 5–30-nm-thick h-BN flakes (a layered electrical insulator free of trapped charges and dangling bonds) were chosen for the top and bottom dielectric[31]. The top and bottom parts were prepared separately using a polymer-based dry transfer technique[32]. For the bottom part, an h-BN flake was picked up on a polymer stamp and placed on the bottom graphite. After dissolving the polymer, Pt metal contacts (about 8 nm) were patterned on the h-BN by standard e-beam lithography, e-beam evaporation and lift-off. For the top part, the top graphite was picked up first, then the top h-BN. Both stacks were then transferred to an oxygen- and water-free glovebox. WTe_2 crystals were exfoliated inside the glovebox and flakes from monolayer to trilayer thickness were optically identified and quickly picked up with the top part; the stack was then completed by transferring onto the lower contacts/h-BN/graphite stack before taking out of the glovebox. Finally, after dissolving the polymer, another step of e-beam lithography and metallization was used to define electrical bonding pads (Au/V) connecting to the metal contacts and the top and bottom gates. Extended Data Fig. 1 shows schematics of the fabrication processes and optical and atomic force microscope (AFM) images of a typical bilayer WTe_2 device (B4).

Estimate of the electric polarization. We use the following simplified model to estimate the spontaneous polarization of the bilayer WTe_2 from the measurements in Fig. 2b. We assume that $d_t = d_b \gg d$, where d is the thickness of the WTe_2, that all conductors (bottom graphite gate, top graphene and bilayer WTe_2) are grounded and have infinite electronic compressibility, and that the areal polarization density P is associated with two thin sheets of areal charge density $\pm P/d$ separated by d. Under these assumptions, when the polarization reverses there is no net flow of charge between the conductors and the WTe_2 remains neutral, and the potential profile between the gates is simply reversed when the polarization flips (Fig. 2d). By Gauss's law

$$\varepsilon_0\varepsilon_{hBN}E_t = \varepsilon_0 E_i + P/d \quad (1)$$

where E_t is the electric field in the h-BN (equal on both sides because the bilayer is neutral) and E_i is the field between the two charge sheets. Because the top graphene and the centre of the bilayer are both at zero potential,

$$2E_t d_t + E_i d = 0 \quad (2)$$

From equations (1) and (2):

$$E_t = \frac{P}{\varepsilon_0(2d_t + \varepsilon_{hBN}d)}$$

The change in E_t when the polarization reverses is then $\delta E_t = 2E_t = 2P/[\varepsilon_0(2d_t + \varepsilon_{hBN}d)]$. With $d_t \approx 10$ nm and $d \approx 1$ nm, the first term in the denominator dominates so $\delta E_t \approx P/(\varepsilon_0 d_t)$ and thus $P \approx \varepsilon_0 d_t \delta E_t = \varepsilon_0 \delta V$. In reality, d_b and d_t can differ by a factor of up to three, the conductors have finite compressibility and the polarization charge is more spread out, which taken together introduce an extra numerical coefficient of order unity.

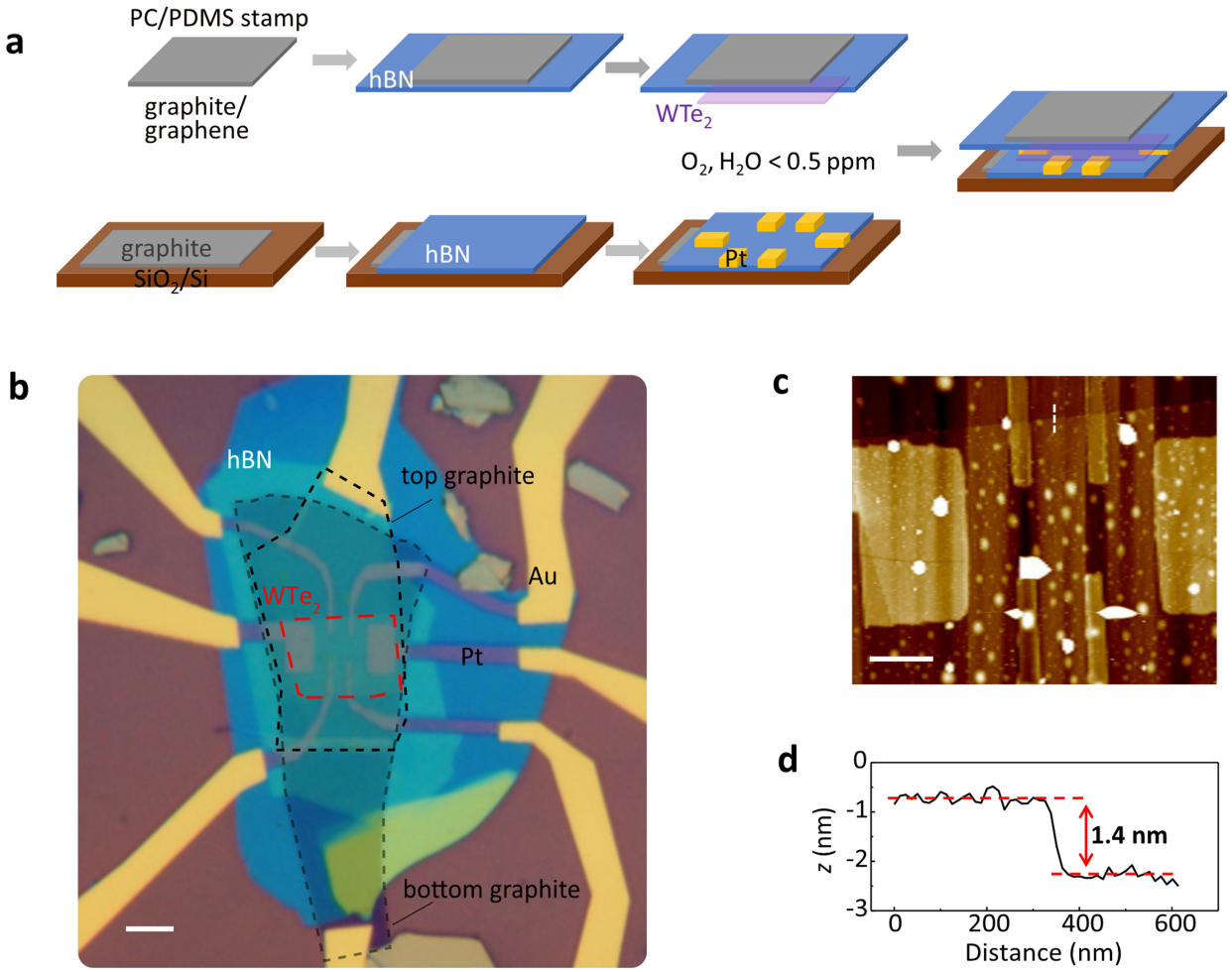
Removing parallel (parasitic) conduction through the graphene in device B2. In device B2 the graphene extends over regions with no WTe_2 underneath so that it acts as a uniform gate for the entire WTe_2 sheet. The quantity that we call G_{gr} is the result of the following measurement, which maximizes sensitivity to only a central region of graphene above the WTe_2. First, we ground two opposing contacts to the graphene and measure only the current that flows from the biased contact to the one opposite, as shown in Fig. 2a. However, because of finite contact resistance, a small portion of this current still flows through graphene not above the WTe_2. To remove this parasitic current component, we set the WTe_2 voltage V_W such that the graphene is at its Dirac-point minimum in the region over the WTe_2. Because the minimum is quite broad, the graphene over the WTe_2 is then insensitive to V_b and the measured dependence on V_b comes from only the parasitic component, which can then be subtracted out. Note that removing it has no effect on the magnitude of the hysteresis.

In Extended Data Fig. 4b, we illustrate this procedure at 220 K. From the inset of Extended Data Fig. 4b, we determine that the graphene above the WTe_2 is at its Dirac point at $V_W = 129$ mV. The red curve shows the conductance of the graphene G_{gr} when $V_W = 129$ mV; the dependence on the back gate is from only the parasitic contribution. Conversely, in Fig. 2b and the blue curve in Extended Data Fig. 4b we measure G_{gr} at $V_W = 0$ mV, at which the graphene is most sensitive to changes in the electric field in the top h-BN E_t, yet also contains the parasitic conductance. The difference between these two curves (at $V_W = 129$ mV and $V_W = 0$ mV) is shown in black. The hysteresis remains, whereas the 'V' shape is mostly removed. The remaining small slope can be explained by the finiteness of the electronic compressibility of the bilayer WTe_2.

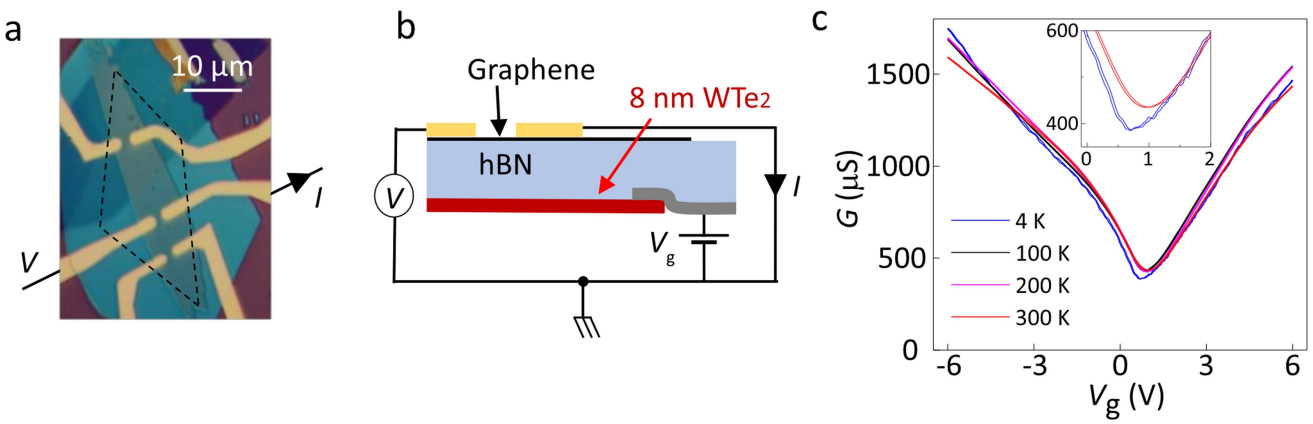
Using Extended Data Fig. 4b we can estimate the ratio of the parasitic current to that flowing above the WTe_2. The area with no WTe_2 has a h-BN thickness of $d_t + d_b = 33$ nm between the graphene and bottom gate. The red curve (with a parasitic V_b dependence) has a maximum slope of $dG_{gr}/dV_b = 17\ \mu S\ V^{-1}$ or $dG_{gr}/dE_t \approx 560\ \mu S\ V^{-1}$ nm after taking into account the h-BN thickness. From the inset curve, using voltage V_W applied to the WTe_2 for gating (with 8 nm h-BN) gives $dG_{gr}/dE_t = 12{,}300\ \mu S\ V^{-1}$ nm. Thus, the parasitic component is only about 5% of the total current.

Data availability. The data presented in this paper and that support the findings of this study are available from the corresponding author on reasonable request.

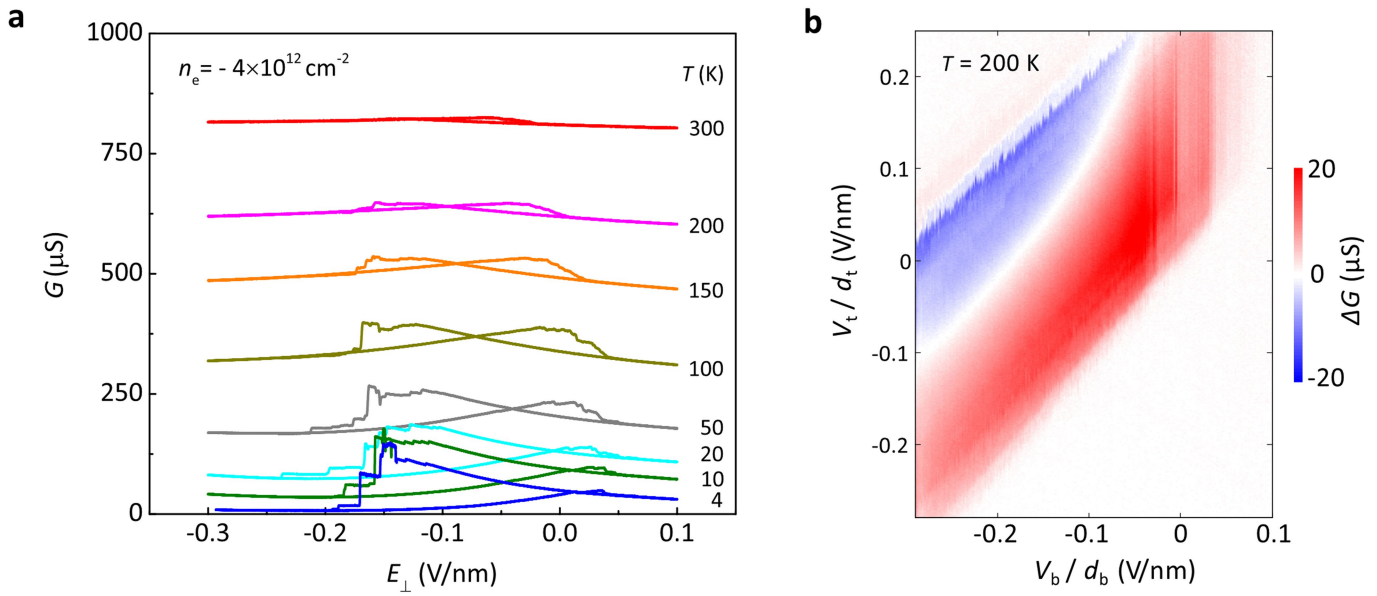
31. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
32. Zomer, P. J., Guimaraes, M. H. D., Brant, J. C., Tombros, N. & van Wees, B. J. Fast pick up technique for high quality heterostructures of bilayer graphene and hexagonal boron nitride. *Appl. Phys. Lett.* **105**, 013101 (2014).



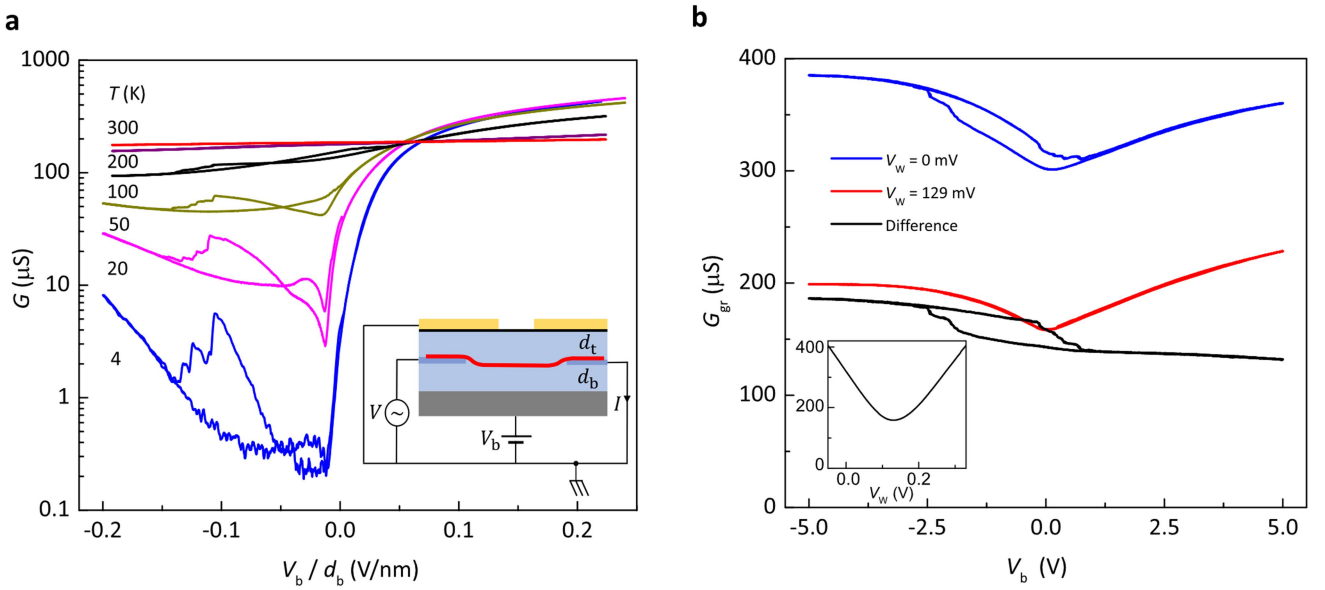
Extended Data Fig. 1 | Bilayer WTe_2 device. a, Essential steps in device fabrication. **b**, Optical image of device B4. The red dashed line outlines the bilayer flake. Scale bar, 5 μm. **c**, AFM topography image of the central region in **b**. Scale bar, 2 μm. **d**, Line cut along the white dashed line in **c**. The step height matches the expected bilayer thickness, about 1.4 nm.



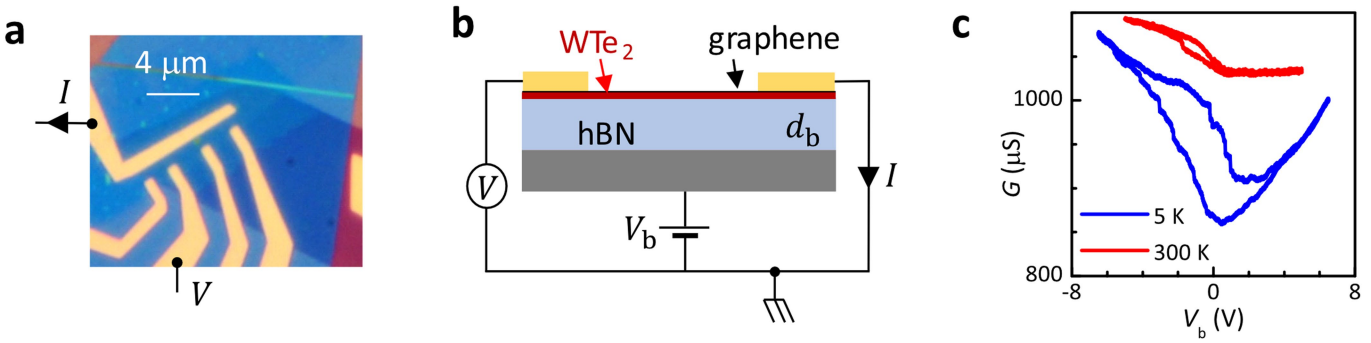
Extended Data Fig. 2 | Thick WTe_2 used as a gate. a, Optical image of device F1, in which a thick (8 nm) WTe_2 flake under 24-nm h-BN is used as a gate for a top graphene sheet. Scale bar, 10 μm. **b**, Schematic cross-section of the device. **c**, Two-terminal conductance G of the graphene as a function of voltage V_g applied to the WTe_2 flake. There is no sign of switching or bistability at any temperature, indicating that no polarization reversal occurs on the WTe_2 surface for fields of up to $E_\perp \approx 0.125$ V nm^{-1}. Inset, close-ups of the graphene Dirac point at 4 K and 300 K.



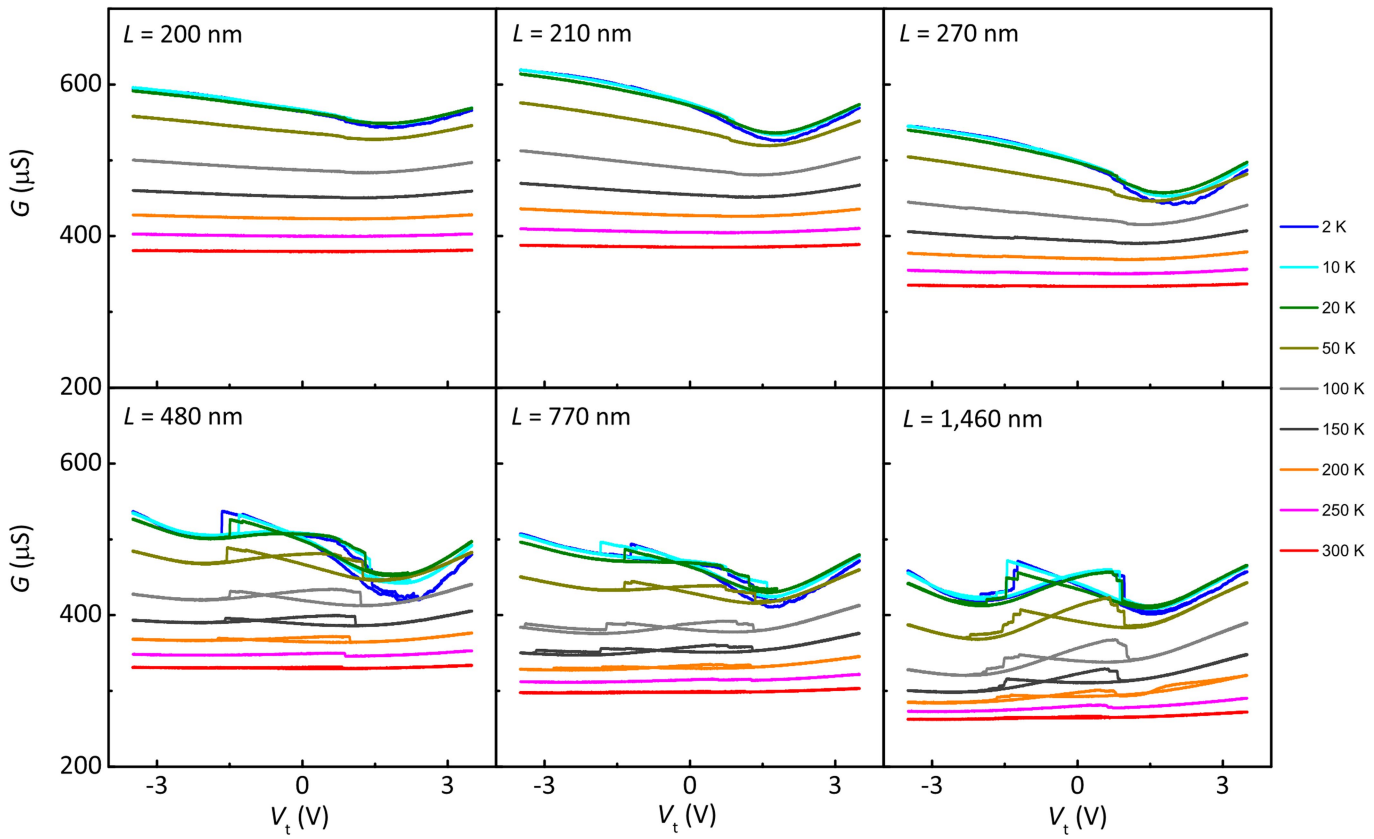
Extended Data Fig. 3 | Switching of an additional bilayer device. **a**, Conductance G versus perpendicular electric field $E_\perp$, at temperatures from 4 K to 300 K and a gate doping level of $n_e = -4 \times 10^{12}$ cm^{-2}, for device B4. **b**, Conductance difference ΔG between the two sweep directions of V_b at 200 K, as plotted in Fig. 3d for device B1.



Extended Data Fig. 4 | Additional transport measurements and removal of parasitic effects in the polarization measurements. a, Conductance G versus V_b for the bilayer WTe_2 in device B2, measured with the top graphene grounded. The hysteresis occurs in exactly the same range of $E_\perp$ as it does in the graphene conductance in Fig. 2b. Note that both n_e and $E_\perp$ change when V_b is swept. The inset shows a schematic configuration of the measurement. **b**, Graphene conductance G_{gr} at 220 K as a function of V_b with the voltage V_W on the bilayer WTe_2 at 0 mV (blue) and 129 mV (red). The black curve is the difference between the blue and red curves. This subtraction removes most of the V_b dependence of the parasitic current that flows through the top graphene, which is not screened from the bottom gate by the WTe_2. Inset, graphene conductance showing the minimum at $V_W = 129$ mV.



Extended Data Fig. 5 | Graphene/bilayer WTe_2 heterostructure showing hysteresis up to room temperature. a, **b**, Device image (**a**) and schematic cross-section (**b**). **c**, The two-terminal conductance *G* shows bistability at both 5 K and room temperature (300 K), implying that the polarization of the WTe_2 is still present in this hybrid structure.



Extended Data Fig. 6 | Length-dependent ferroelectric behaviour in trilayer WTe_2 for temperatures from 2 K to 300 K. All measurements are performed at $V_b = 0$ in two-terminal configurations, where the contact separation ranges from 200 nm to 1,490 nm. For all devices mentioned above and in the main text, the contacts are separated by 1–2 μm. However, if we reduce the contact separation to a few hundred nanometers (270 nm), the metal contacts prevent the polarization from switching. For a contact separation (L) of more than 480 nm, the transfer characteristics show similar hysteric behaviour as in Fig. 1c, d and Extended Data Fig. 3a. Because V_b is always grounded, $E_\perp$ and n_e change simultaneously as we sweep V_t.

Extended Data Table 1 | Thickness of h-BN dielectrics and corresponding areal capacitances for the WTe_2 devices

Device label	WTe_2	top hBN (nm)	bottom hBN (nm)	C_t (1×10^{-3} F/m^2)	C_b (1×10^{-3} F/m^2)
M1	monolayer	6	28	5.9	1.3
B1	bilayer	12	20	3.0	1.8
B2	bilayer	8	25	4.4	1.4
B3	bilayer	NA	24	NA	1.5
B4	bilayer	10	21	3.5	1.7
T1	trilayer	5.5	23	6.4	1.1*
F1	8 nm	24	NA	1.5	NA

We define the gate-induced density imbalance to be $n_e = (C_t V_t + C_b V_b)/e$ and $E_\perp = (-C_t V_t + C_b V_b)/(2\varepsilon_{h\text{-}BN}\varepsilon_0)$, where the geometric areal capacitances are $C_t = \varepsilon_{hBN}\varepsilon_0/d_t$ and $C_b = \varepsilon_{hBN}\varepsilon_0/d_b$, $\varepsilon_{hBN} \approx 4$ is the dielectric constant of h-BN, and d_t and d_b are the thicknesses of the top and bottom h-BN flakes, respectively. All thicknesses were obtained from AFM images. In device B3, the WTe_2 flake is directly under the top graphene (no top h-BN). In device F1, the thick WTe_2 is directly on the bottom graphite (no bottom h-BN).
*For device T1, there is no bottom graphite; instead, we used the metallic silicon substrate as the bottom gate, the areal capacitance then being $C_b = \varepsilon_0/(d_b/\varepsilon_{hBN} + d_{SiO2}/\varepsilon_{SiO2})$. We did not make a four-layer device so we do not know at exactly what thickness polarization switching ceases to be possible.

LETTER

https://doi.org/10.1038/s41586-018-0357-y

Room-temperature electrical control of exciton flux in a van der Waals heterostructure

Dmitrii Unuchek[1,2,4], Alberto Ciarrocchi[1,2,4], Ahmet Avsar[1,2], Kenji Watanabe[3], Takashi Taniguchi[3] & Andras Kis[1,2]*

Devices that rely on the manipulation of excitons—bound pairs of electrons and holes—hold great promise for realizing efficient interconnects between optical data transmission and electrical processing systems. Although exciton-based transistor actions have been demonstrated successfully in bulk semiconductor-based coupled quantum wells[1–3], the low temperature required for their operation limits their practical application. The recent emergence of two-dimensional semiconductors with large exciton binding energies[4,5] may lead to excitonic devices and circuits that operate at room temperature. Whereas individual two-dimensional materials have short exciton diffusion lengths, the spatial separation of electrons and holes in different layers in heterostructures could help to overcome this limitation and enable room-temperature operation of mesoscale devices[6–8]. Here we report excitonic devices made of MoS_2–WSe_2 van der Waals heterostructures encapsulated in hexagonal boron nitride that demonstrate electrically controlled transistor actions at room temperature. The long-lived nature of the interlayer excitons in our device results in them diffusing over a distance of five micrometres. Within our device, we further demonstrate the ability to manipulate exciton dynamics by creating electrically reconfigurable confining and repulsive potentials for the exciton flux. Our results make a strong case for integrating two-dimensional materials in future excitonic devices to enable operation at room temperature.

Solid-state devices use particles and their quantum numbers for their operation, with electronics being the ubiquitous example. The need to improve power efficiency of charge-based devices and circuits is motivating research into new devices that would rely on other principles. Candidates so far include spintronics and photonics[9,10]. Excitons—electrically neutral quasi-particles formed by bound electrons and holes—can also be manipulated in solid-state systems. The development of such excitonic devices has so far been hindered by the absence of a suitable system that would enable room-temperature manipulation of excitons, limiting the expansion of the field. Here, we demonstrate room-temperature excitonic devices based on atomically thin semiconductors. These devices could open the way for wider studies and applications of excitonic devices in the academic and industrial sectors[11]. Many applications can be envisaged, because excitons could be used to efficiently couple optical data transmission and electronic processing systems. Although fast optical switches have already been

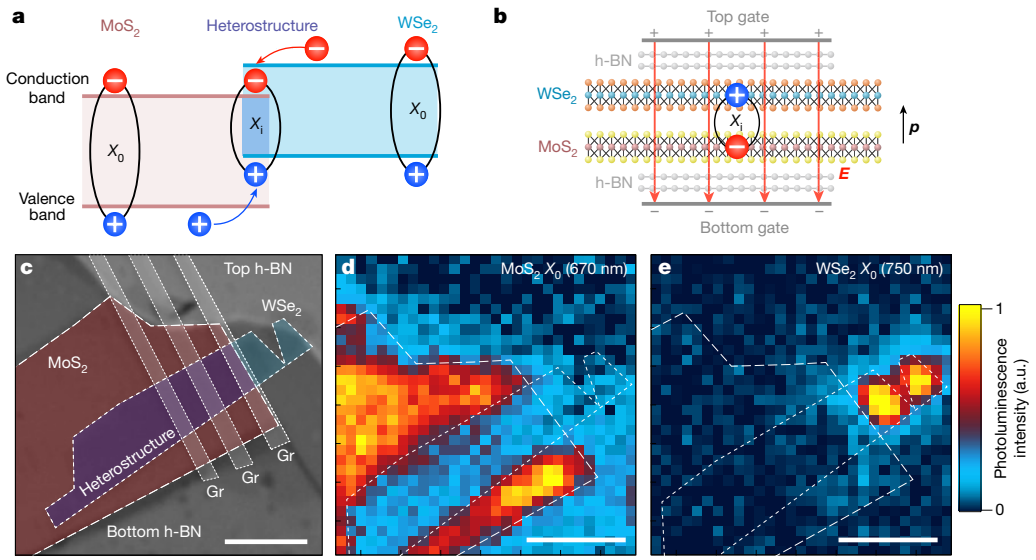


Fig. 1 | Interlayer excitons in the WSe_2–MoS_2 van der Waals heterostructure. a, Type-II band alignment in the WSe_2–MoS_2 heterostructure with intralayer (X_0) and interlayer (X_i) excitons. The red and blue areas represent the bands in the two materials and the heterobilayer. Positive and negative symbols indicate holes and electrons, respectively. **b**, Schematic depiction of the WSe_2–MoS_2 heterostructure, showing the heterobilayer encapsulated in hexagonal boron nitride (h-BN) and the top and bottom gates. The interlayer exciton has a permanent out-of-plane dipole moment $\boldsymbol{p}$ that allows manipulation via the electric field $\boldsymbol{E}$. **c**, False-colour optical image of the device, highlighting the different materials. **d**, **e**, Spatial maps of photoluminescence at 670 nm (**d**) and 750 nm (**e**), corresponding to MoS_2 and WSe_2 intralayer excitonic resonances, respectively. Photoluminescence is quenched in the heterostructure area owing to efficient charge transfer. Scale bars, 5 μm. a.u., arbitrary units.

[1]Electrical Engineering Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [2]Institute of Materials Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [3]National Institute for Materials Science, Tsukuba, Japan. [4]These authors contributed equally: Dmitrii Unuchek, Alberto Ciarrocchi. *e-mail: andras.kis@epfl.ch

demonstrated[12,13], the comparably large size (about 10 μm)[14,15] of such devices limits packing density. This can be overcome in excitonic devices, the characteristic size of which is determined by that of electronic field-effect transistors (FETs).

Owing to their finite binding energy E_b, excitons can exist up to temperatures of around $T \propto E_b/k_B$, where k_B is the Boltzmann constant. In a conventional III–V-semiconductor coupled quantum well with a size of a few nanometres, the relatively small binding energy of around 10 meV permits the observation of excitons only at cryogenic temperatures (less than 100 K)[3]. To reach higher temperatures, different materials are required. To this end, systems with higher E_b (in the range of tens of millielectronvolts) have been explored more recently, such as (Al,Ga)N/GaN (ref. [16]) or ZnO (ref. [17]). Two-dimensional semiconductors such as transition-metal dichalcogenides have even larger exciton binding energies, which can exceed 500 meV in some cases owing to strong quantum confinement[4,5]. This could enable the realization of excitonic devices that operate at room temperature[18].

Although intralayer excitons have relatively short lifetimes (about 10 ps)[7,19], the spatial separation of holes and electrons in interlayer excitons results in lifetimes more than two orders of magnitude longer, well in the nanosecond range[6]. For the device presented here, we take advantage of interlayer excitons in an atomically thin MoS_2–WSe_2 heterostructure. Type-II band alignment[20,21] (Fig. 1a) results in charge separation between the constituent materials, with electrons and holes residing in MoS_2 and WSe_2, respectively. The formation of indirect excitons is marked by the appearance of a new photoluminescence emission peak[22], redshifted by about 75 meV with respect to the intralayer exciton of the WSe_2 monolayer. In Extended Data Fig. 1b we present a typical photoluminescence spectrum obtained from such a heterostructure on SiO_2, in which the spectral signature of the interlayer exciton is clearly visible (dark blue line), together with the individual WSe_2 and MoS_2 monolayers (blue and red lines, respectively). Recent reports[23] suggest that excitons in the MoS_2–WSe_2 system are not only spatially indirect, but also momentum-indirect owing to lattice mismatch. The phonon-assisted nature of the emission process further reduces the exciton recombination rate, yielding a longer lifetime[8,24]. Such an extended lifetime can be used to obtain interlayer exciton diffusion over a scale of micrometres, even at room temperature.

To obtain a pristine surface, the heterostructure is encapsulated in hexagonal boron nitride and annealed in high vacuum. Multiple transparent top gates are fabricated out of few-layer graphene. This double-gate configuration allows us to apply a vertical electric field without changing the carrier concentration in the MoS_2–WSe_2 heterostructure. In Fig. 1c we show a false-colour optical micrograph of the resulting stack. We characterize the structure by using photoluminescence mapping at room temperature, under 647-nm excitation. In Fig. 1d, e and Extended Data Fig. 1 we show the intralayer emission distribution at the wavelengths characteristic of MoS_2 (670 nm), WSe_2 (760 nm) and the interlayer exciton (785 nm). Whereas individual monolayers appear to be homogeneously bright, emission from the heterostructure region is uniformly quenched by more than three orders of magnitude, owing to the efficient charge transfer between layers[24]. Even with this strong quenching, we are able to detect the interlayer peak in the photoluminescence spectra (Extended Data Fig. 2), confirming the generation of interlayer excitons. Because this effect has a central role in our work, we fabricated three more heterostructures encapsulated in hexagonal boron nitride, confirming the reproducibility of this result (Extended Data Fig. 3).

Given that excitons do not carry a net electric charge, we do not expect their flow to be influenced by the direct application of an in-plane electric field. However, the confinement of oppositely charged carriers in different layers results in a well-defined interlayer-exciton dipole moment $\boldsymbol{p}$ with an out-of-plane (z) direction (Fig. 1b). An electric field $E_z(x, y)$ perpendicular to the crystal plane can then be used to shift the exciton energy by $\delta E = -p_z E_z$, while its lateral modulation drives the exciton motion towards regions of lower energy. Exciton

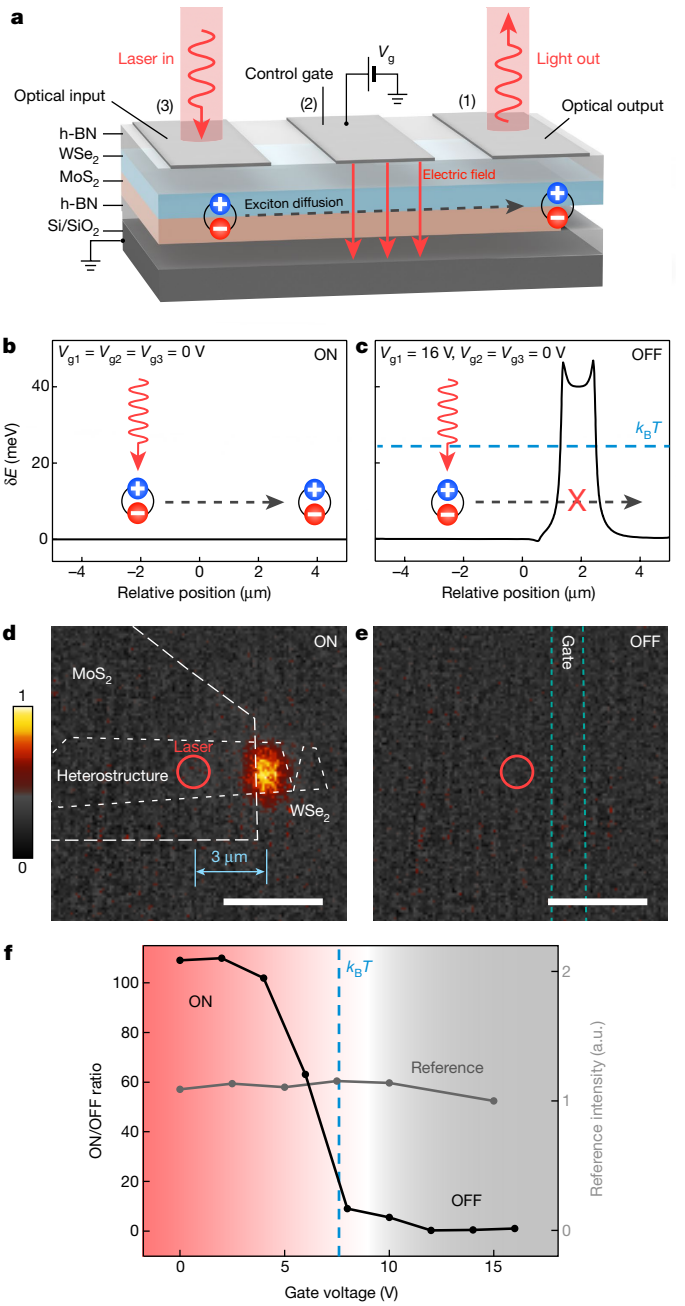


Fig. 2 | Excitonic transistor operation at room temperature. a, The application of gate voltages (V_{g1}, V_{g2}, V_{g3}) to transparent graphene electrodes (gates 1–3) can engineer a potential landscape for the diffusion of excitons, controlling their flux through the device. **b**, **c**, Calculated energy variation δE for the excitons in the ON (free diffusion; **b**) and OFF (potential barrier; **c**) states. Red arrows represent laser excitation; the bound charges and black dashed arrows denote the excitons and their diffusion, respectively. **d**, **e**, Corresponding images of exciton emission. Dashed lines indicate the positions of the different layers that form the heterostructure and the top graphene gate (gate 1). The laser spot is represented by the red circle. Colour scale indicates the normalized photoluminescence intensity. Scale bars, 5 μm. **f**, Gate dependence of the ON/OFF ratio for optical excitation 3 μm away from the emission centre (left axis). The right axis shows the reference data, which were acquired with the incident laser beam located directly on the emission centre (input–output distance, $d_{i-o} = 0$ μm). The measured emission intensity is normalized by the OFF-state value at $V_{g1} = 15$ V. The background shading indicates the ON (red) and OFF (grey) states. The blue dashed line represents the gate voltage at which the barrier height is equal to the thermal energy.

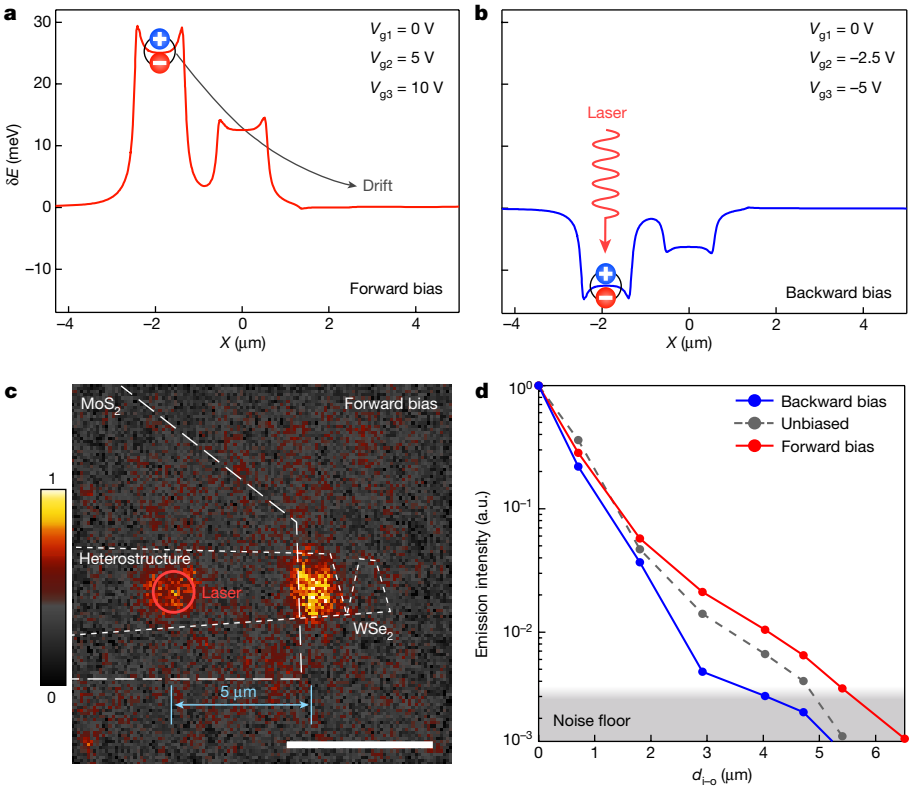


Fig. 3 | Biasing of the excitonic device. a, **b**, Calculated energy profile δE of the indirect exciton as a function of lateral coordinate X for the forward (**a**) and backward (**b**) bias cases. The black solid line indicates the direction of exciton drift. **c**, Image showing exciton emission from the device when injecting at a distance $d_{i-o} = 5\,\mu m$ from the emission area. Colour scale, dashed lines and red circles as in Fig. 2d, e. Scale bar, 5 μm. **d**, Normalized output intensity as a function of the distance d_{i-o} between optical injection and the emission point, for the forward (red) and backward (blue) bias configurations, compared to the unbiased case (grey). The grey shading indicates the noise floor. Exciton diffusion over a distance of 5.5 μm is achieved.

dynamics in the longitudinal direction can be modelled by a diffusion equation with an external potential (see Methods):

$$D\frac{\partial^2 n}{\partial x^2} + \frac{D}{k_B T}\frac{\partial}{\partial x}\left(n\frac{\partial \varphi}{\partial x}\right) + G - \frac{n}{\tau} = \frac{\partial n}{\partial t} \quad (1)$$

where n, D and τ are the interlayer-exciton concentration, diffusion coefficient and lifetime, respectively, φ is the exciton potential (including the electrostatic contribution $\varphi_{el} = -p_z E_z$) and G is the optical generation rate. This simple model qualitatively shows how the application of an electric field E_z can affect interlayer exciton diffusion, as we discuss later.

We first demonstrate an electrically controlled excitonic switch, represented schematically in Fig. 2a. Laser light focused inside the heterostructure area (input) generates interlayer excitons, which diffuse along the channel of the heterostructure. However, the low brightness of interlayer emission makes monitoring the operation of the device challenging. For this reason, we use the exposed WSe_2 that extends out of the heterostructure as a bright emitter. Here, interlayer excitons diffuse towards the edge of the heterostructure. During this diffusion process, interlayer excitons are expected to dissociate into single carriers, which are allowed to diffuse inside monolayer MoS_2 and WSe_2, where they experience recombination with native charges, resulting in bright emission. The emitted radiation is recorded simultaneously using a charge-coupled device (CCD) camera and a spectrometer (see Methods), to obtain spatial and spectral emission profiles. This allows us to further confirm the presence and diffusion of interlayer excitons inside the heterobilayer (Extended Data Fig. 2). In the absence of applied fields (Fig. 2b), excitons diffuse away from the pumping area (red circle in Fig. 2d), owing to temperature and concentration gradients[25–27], and reach the recombination site, approximately 3 μm away. Comparison of pumping and emission profiles (Extended Data Fig. 4) lets us exclude the possibility of a direct excitation of monolayer WSe_2 by the low-intensity tail of the laser spot. This situation (bright output) is shown in the emission image in Fig. 2d and corresponds to the ON state of the excitonic transistor. On the contrary, by introducing a potential barrier higher than $k_B T$ on the path of the diffusing excitons (Fig. 2c), we impede their motion, resulting in the suppression of light emission (Fig. 2e). In this way, we can achieve efficient electrical modulation of the output emission, as shown in Fig. 2f, in which the emission intensity (normalized by the value in the OFF state, corresponding to $V_{g1} = +16$ V) is plotted as a function of applied voltage. For reference, we also plot the intensity modulation observed when the laser beam is located on the emission centre (input–output distance $d_{i-o} = 0\,\mu m$). The switching threshold is around 8 V, which corresponds well with the calculated exciton energy modulation of $\delta E \approx k_B T \approx 25$ meV (blue dashed line in Fig. 2f). This result is consistent with our model: because the height of the energy barrier starts to become comparable to thermal excitation, it is now possible to block the diffusion of exciton flux. We extract an intensity ON/OFF ratio larger than 100, limited by the noise level of the set-up in the OFF state (see also Extended Data Figs. 4, 5). Such a high ratio results from the realization of an excitonic transistor with complete suppression of emission in the OFF state. This effect is also clearly visible in the spectrum of the emitted light, in which the WSe_2 peak is selectively suppressed when the device is in the OFF state (Extended Data Fig. 6). We also note that strong emission from MoS_2 is detected in both states, because excitons can diffuse freely in other directions.

An alternative mechanism that could in principle explain the recombination far away from the excitation spot is based on the diffusion of single carriers rather than interlayer excitons. It has been shown that such carriers (holes in particular) can have long lifetimes[6,28,29].

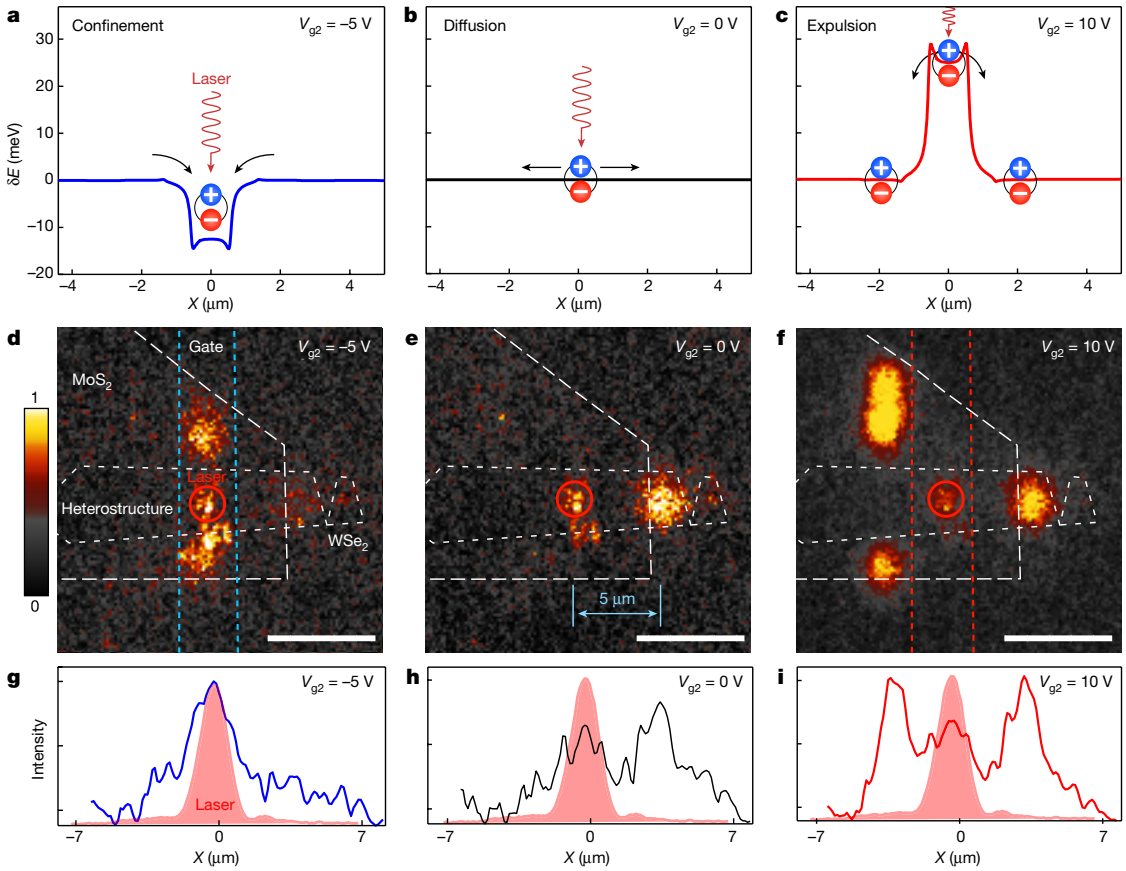


Fig. 4 | Electrically reconfigurable energy landscape. a–c, Calculated energy profile δE of the indirect exciton for the cases of a potential well (**a**), free diffusion (**b**) and a potential barrier (**c**). **d–f**, Imaging of exciton emission for the configurations shown in **a–c**. Incident laser light (red circle) is focused on top of gate 2. Dashed lines indicate positions of different layers that form the heterostructure and the graphene top gate 2; colour scale as in Fig. 2d, e. Scale bars, 5 μm. **g–i**, Cross-section of the intensity profile along the device channel, integrated over its width, for the three configurations. The red-shaded underlay represents the profile of the excitation laser.

However, experimental observations indicate that this is not the dominant mechanism in our heterostructure. First, we observe the production of interlayer excitons directly in the excitation area, even if the intensity is low. Second, for a flux of single carriers, the voltage modulation necessary to counteract thermal excitation and block the single-particle flux would be about 50 mV, more than two orders of magnitude lower that the gate voltage of approximately 8 V required in our experimental result shown in Fig. 2. Finally, this mechanism would also result in different emission profiles for different regimes of device operation (see Extended Data Fig. 7).

To exclude the possibility that the observed effect arises from an unwanted modulation of the charge carrier density in WSe_2, we perform a calibration experiment in which the excitation light is focused on the output area ($d_{i-o}=0$) and the device is biased as before. This reference experiment is discussed in detail in Methods and the result is presented in Fig. 2f (grey curve); it shows that only a comparably small modulation of WSe_2 emission intensity is observed. This confirms that the energy barrier is the origin of the switching behaviour. We study the dependence of the ON/OFF ratio on d_{i-o} further (Extended Data Fig. 8) by keeping the voltage profile constant and optically injecting excitons at different distances from the output point. Consistent with our model, we observe efficient modulation when the laser is focused beyond the energy barrier, with emission intensity decreasing with increasing d_{i-o} owing to long-distance diffusion. The diffusion length can be doubled at lower temperature (4.7 K), resulting in operation over a longer distance (Extended Data Fig. 9).

Having demonstrated that we can block or allow spontaneous exciton diffusion, we go further by creating a drift field in the desired direction, in analogy with the source–drain bias of a conventional FET. We show this type of operation in Fig. 3, with all three electrodes used to create a potential ladder going upwards or downwards with respect to the excitation point (Fig. 3a, b). When excitons encounter a gradually decreasing energy profile (forward bias), their diffusion is enhanced by a drift term, allowing us to operate the device with a larger distance between optical input and output. As shown in Fig. 3c, this regime of electrically assisted diffusion can result in exciton transport over a distance of 5 μm. To obtain a more quantitative estimate of the induced modulation, we measure the dependence of the emission intensity on the distance from the laser spot as it is displaced away from the output area at fixed gate voltages. The results (Fig. 3d) show that the length over which excitons diffuse can be effectively modulated from 5.5 μm to 3 μm, compared to about 4.5 μm in the unbiased case. The modulation of the effective diffusion length with the potential φ_{el} qualitatively follows the model introduced in equation (1).

We further use the multi-gate configuration to demonstrate more complex and electrically reconfigurable types of potential landscape and related device operation. In Fig. 4a–c we present the energy profiles calculated for free diffusion (Fig. 4b) compared with a potential well (Fig. 4a) and a repulsive barrier (Fig. 4c) produced by the central gate (gate 2), while the side gates (1 and 3) are kept grounded. In this case, the position of the optical pump is centred on the middle electrode, which corresponds to the centre of the well or barrier. In Fig. 4d, g we show the CCD camera image and related emitted intensity profile along the device channel for the case of the potential well. We observe photoluminescence emission only from the narrow area below the central contact, which is indicative of electrical confinement of the excitonic cloud. Conversely, when applying a positive voltage to create a 'potential hill' (Fig. 4f, i), we see an expulsion of excitons from the pumping area

with the appearance of bright emission spots outside the middle section of the device, owing to excitons drifting along the energy profile and recombining on the edges of the heterostructure. This is evident from a comparison with the free-diffusion case in Fig. 4e, h. Interestingly, we also observe higher-energy emission from the neighbouring MoS_2 monolayer parts inside the well in the case of exciton confinement. A similar effect is also observed during exciton expulsion, with bright spots appearing at the edges of the heterostructure around the repulsive potential. Further inspection of the emission spectra from Fig. 4d, f confirms this, with the intensity of monolayer peaks decreasing (increasing) when confining (anti-confining) the excitons (Extended Data Fig. 6). As also discussed in Methods, the observed MoS_2 emission is affected by the local inhomogeneity of the substrate and by the optical filters used. As discussed earlier, the diffusion of single particles and their recombination with native charges that are available in the monolayers could have a role in light emission that extends from the edges of the heterobilayer into the monolayers.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0357-y.

Received: 29 November 2017; Accepted: 4 June 2018;
Published online 25 July 2018.

1. High, A. A., Hammack, A. T., Butov, L. V., Hanson, M. & Gossard, A. C. Exciton optoelectronic transistor. *Opt. Lett.* **32**, 2466–2468 (2007).
2. High, A. A., Novitskaya, E. E., Butov, L. V., Hanson, M. & Gossard, A. C. Control of exciton fluxes in an excitonic integrated circuit. *Science* **321**, 229–231 (2008).
3. Grosso, G. et al. Excitonic switches operating at around 100 K. *Nat. Photon.* **3**, 577–580 (2009).
4. Cheiwchanchamnangij, T. & Lambrecht, W. R. L. Quasiparticle band structure calculation of monolayer, bilayer, and bulk MoS_2. *Phys. Rev. B* **85**, 205302 (2012).
5. He, K. et al. Tightly bound excitons in monolayer WSe_2. *Phys. Rev. Lett.* **113**, 026803 (2014).
6. Rivera, P. et al. Observation of long-lived interlayer excitons in monolayer $MoSe_2$–WSe_2 heterostructures. *Nat. Commun.* **6**, 6242 (2015).
7. Palummo, M., Bernardi, M. & Grossman, J. C. Exciton radiative lifetimes in two-dimensional transition metal dichalcogenides. *Nano Lett.* **15**, 2794–2800 (2015).
8. Miller, B. et al. Long-lived direct and indirect interlayer excitons in van der Waals heterostructures. *Nano Lett.* **17**, 5229–5237 (2017).
9. Žutić, I., Fabian, J. & Das Sarma, S. Spintronics: fundamentals and applications. *Rev. Mod. Phys.* **76**, 323–410 (2004).
10. Miller, D. A. B. Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE* **88**, 728–749 (2000).
11. Butov, L. V. Excitonic devices. *Superlattices Microstruct.* **108**, 2–26 (2017).
12. Liu, A. et al. A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor. *Nature* **427**, 615–618 (2004).
13. Melikyan, A. et al. High-speed plasmonic phase modulators. *Nat. Photon.* **8**, 229–233 (2014).
14. Xu, Q., Schmidt, B., Pradhan, S. & Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **435**, 325–327 (2005).
15. Schmidt, B., Xu, Q., Shakya, J., Manipatruni, S. & Lipson, M. Compact electro-optic modulator on silicon-on-insulator substrates using cavities with ultra-small modal volumes. *Opt. Express* **15**, 3140–3148 (2007).
16. Fedichkin, F. et al. Room-temperature transport of indirect excitons in AlGaN/GaN quantum wells. *Phys. Rev. Appl.* **6**, 014011 (2016).
17. Kuznetsova, Y. Y. et al. Transport of indirect excitons in ZnO quantum wells. *Opt. Lett.* **40**, 3667–3670 (2015).
18. Fogler, M. M., Butov, L. V. & Novoselov, K. S. High-temperature superfluidity with indirect excitons in van der Waals heterostructures. *Nat. Commun.* **5**, 4555 (2014).
19. Robert, C. et al. Exciton radiative lifetime in transition metal dichalcogenide monolayers. *Phys. Rev. B* **93**, 205423 (2016).
20. Kang, J., Tongay, S., Zhou, J., Li, J. & Wu, J. Band offsets and heterostructures of two-dimensional semiconductors. *Appl. Phys. Lett.* **102**, 012111 (2013).
21. Chiu, M.-H. et al. Determination of band alignment in the single-layer MoS_2/WSe_2 heterojunction. *Nat. Commun.* **6**, 7666 (2015).
22. Fang, H. et al. Strong interlayer coupling in van der Waals heterostructures built from single-layer chalcogenides. *Proc. Natl Acad. Sci. USA* **111**, 6198–6202 (2014).
23. Kunstmann, J. et al. Momentum-space indirect interlayer excitons in transition-metal dichalcogenide van der Waals heterostructures. *Nat. Phys.* https://doi.org/10.1038/s41567-018-0123-y (2018).
24. Kim, J. et al. Observation of ultralong valley lifetime in WSe_2/MoS2 heterostructures. *Sci. Adv.* **3**, e1700518 (2017).
25. Cadiz, F. et al. Exciton diffusion in WSe_2 monolayers embedded in a van der Waals heterostructure. *Appl. Phys. Lett.* **112**, 152106 (2018).
26. Kulig, M. et al. Exciton diffusion and halo effects in monolayer semiconductors. *Phys. Rev. Lett.* **120**, 207401 (2018).
27. Onga, M., Zhang, Y., Ideue, T. & Iwasa, Y. Exciton Hall effect in monolayer MoS_2. *Nat. Mater.* **16**, 1193–1197 (2017).
28. Nagler, P. et al. Interlayer exciton dynamics in a dichalcogenide monolayer heterostructure. *2D Mater.* **4**, 025112 (2017).
29. Wang, R. et al. Ultrafast and spatially resolved studies of charge carriers in atomically thin molybdenum disulfide. *Phys. Rev. B* **86**, 045406 (2012).

Acknowledgements We are grateful to K. Marinov and D. Ovchinnikov for discussion. We acknowledge the help of Z. Benes (CMI) with electron-beam lithography. D.U., A.C., A.A. and A.K. acknowledge support by the Swiss National Science Foundation (grant 153298), H2020 European Research Council (ERC, grant 682332) and Marie Curie-Sklodowska-Curie Actions (COFUND grant 665667). A.K. acknowledges funding from the European Union's Horizon H2020 Future and Emerging Technologies under grant agreement number 696656 (Graphene Flagship). K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the MEXT, Japan and JSPS KAKENHI grant numbers JP15K21722 and JP25106006.

Reviewer information *Nature* thanks W. Gao, A. Tartakovskii and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.C. and A.A. fabricated the devices, D.U. and A.C. performed the optical measurements and analysed the data. K.W. and T.T. grew the h-BN crystals. A.K. initiated and supervised the project. D.U., A.C., A.A. and A.K. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0357-y.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.K.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Device fabrication. The heterostructure was fabricated using polymer-assisted transfer (see Extended Data Fig. 10) of flakes of hexagonal boron nitride (h-BN), WSe_2 (HQ Graphene) and MoS_2 (SPI). Flakes were first exfoliated on a polymer double layer, as described previously[30]. Once monolayers were optically identified, the bottom layer was dissolved with a solvent and free-floating films with flakes were obtained. These were transferred using a custom-built set-up with micromanipulators to carefully align flakes on top of each other. During the transfer process, the sharp edges of the flakes were aligned to obtain a twist angle between the two crystal axes close to 0° (or 60°). However, in the case of MoS_2–WSe_2 heterobilayers, the alignment has been shown to be not critical for the observation of interlayer excitons[23,31]. This is due to the indirect (in reciprocal space) nature of the transition and to the considerable lattice mismatch between the two layers (about 4%). Polymer residue was removed with a hot acetone bath. Once completed, the stack was thermally annealed in high vacuum at 10^{-6} mbar for 6 h. Few-layer graphene flakes were obtained by exfoliation from graphite (NGS) on Si/SiO_2 substrates and patterned in the desired shape by electron-beam lithography and oxygen plasma etching. After thermal annealing, the patterned flakes were transferred on top of the van der Waals stack using a polymer-assisted transfer and the entire structure was annealed again in high vacuum. Finally, electrical contacts were fabricated by electron-beam lithography and metallization (60 nm/2 nm Au/Ti).

Optical measurements. All measurements presented here were performed in vacuum at room temperature unless specified otherwise. Excitons were optically pumped by a continuous-wave 647-nm laser diode focused to the diffraction limit with a beam size of about 1 μm. The incident power was 250 μW. The spectral and spatial characteristics of the device emission were analysed simultaneously. The emitted light was acquired using a spectrometer (Andor) and the laser line was removed with a long-pass 650-nm edge filter. For spatial imaging, we used a long-pass 700-nm edge filter so that the laser light and most of the MoS_2 emission were blocked. Filtered light was acquired by a CCD camera (Andor Ixon). The room-temperature photoluminescence spectrum of MoS_2 shown in Extended Data Fig. 1b was obtained under 150-μW excitation at 647 nm, whereas monolayer WSe_2 and the heterostructure fabricated on SiO_2 substrate were characterized under 488-nm excitation.

Owing to the small separation between the interlayer and the intralayer WSe_2 exciton peaks, it is not possible to completely distinguish them in the images acquired on the CCD. The tail of the WSe_2 monolayer peak normally overlaps with the spectral line of the interlayer exciton considerably, meaning that weak luminescence around 785 nm can be observed even on monolayer WSe_2 (Extended Data Fig. 3e), which is not due to interlayer excitons.

Because of the use of the 700-nm filter, the emission from monolayer MoS_2 is in principle not observable on the CCD. However, some light can be transmitted when the broadening of the photoluminescence peak results in a low-energy tail (see Extended Data Fig. 11) extending beyond 700 nm. Local inhomogeneity in the substrate can affect this broadening, which could explain why the observed MoS_2 luminescence in Fig. 4f comes mostly from the left part of the device.

Low-temperature measurements (Extended Data Fig. 9) were performed in a liquid-helium, continuous-flow cryostat (Oxford Instruments).

Reference experiment. We performed a reference experiment to exclude spurious effects that could compromise the interpretation of the data. First, we observed how the photoluminescence emission from monolayer WSe_2 changes when gating the device using the back gate. For this purpose, we excited the exposed WSe_2 with the laser beam directly and recorded the photoluminescence spectra. When applying voltage to the back gate, a modulation in the emission intensity is clearly observable (Extended Data Fig. 12a). We repeated the same measurement, but instead of applying a voltage between the flake and the back gate, we biased the top and back gates, thus generating a vertical electric field inside the device. In this case, we cannot observe any substantial change in the emission intensity (Extended Data Fig. 12b). This allows us to rule out the possibility that the switching action that we observe could be due to a suppression of photoluminescence from a changing doping level in the material.

Image processing. To aid the interpretation of images from the CCD camera, we performed several image-processing steps using ImageJ[32]. We first subtracted from the original image a background image obtained without laser illumination, to account for ambient light noise. In some cases, a simple background was not sufficient to compensate for the presence of spurious signals from unwanted reflections or changing ambient background. In these cases, a background image was generated by applying the rolling-ball algorithm in ImageJ. Contrast was adjusted to cover the range of values in the image. We provide an example of the procedure in Extended Data Fig. 13.

Modelling exciton diffusion. The dynamics of the exciton in the channel of our device can be modelled by one-dimensional diffusion in the presence of an external potential $\varphi(x)$ (temperature, electrostatic potential or dipole–dipole interaction). The gradient of exciton concentration $n(x)$ drives diffusion current j_{diff} while the potential gradient causes drift j_{drift}:

$$j_{\mathrm{diff}} = -D\frac{\partial n}{\partial x},\ j_{\mathrm{drift}} = -\mu n\frac{\partial \varphi}{\partial x}$$

where μ is the exciton mobility, which is related to the diffusion coefficient D and the thermal energy k_BT by the Einstein relation $D=\mu k_BT$. We also include an exciton generation rate G by means of optical pumping and an exciton recombination rate R, which is related to the exciton lifetime as $R=-n/\tau$. From the exciton continuity equation we then obtain equation (1).

In our system, in which excitons have a built-in vertical dipole moment $\boldsymbol{p}$, the electrostatic potential induced by the vertical electric field is $\varphi_{\mathrm{el}}=-E_zp_z$. Because we use continuous-wave excitation, we assume a steady-state case ($\partial n/\partial t=0$). Considering φ_{el} as the main contribution to exciton drift, we obtain

$$D\frac{\partial^2 n}{\partial x^2}-\frac{Dp_z}{k_BT}\frac{\partial}{\partial x}\left(n\frac{\partial E_z}{\partial x}\right)+G-\frac{n}{\tau}=0$$

We simplify the model further by assuming two fundamentally different regions, shown in Extended Data Fig. 14. First region is under constant homogeneous excitation so that the concentration reaches an equilibrium value with equal recombination and generation rates ($R+G=0$). The equilibrium concentration is then $n_0=G\tau$. Outside of the pumping region, excitons diffuse away, driven by the concentration and potential gradients:

$$D\frac{\partial^2 n}{\partial x^2}-\frac{Dp_z}{k_BT}\frac{\partial}{\partial x}\left(n\frac{\partial E_z}{\partial x}\right)-\frac{n}{\tau}=0$$

The case of diffusion in the absence of an external field can be solved analytically, revealing exponential decay of exciton density from the pumping region with a characteristic distance that corresponds to the diffusion length $l_{\mathrm{diff}}=\sqrt{D\tau}$: $n_{\mathrm{free}}(x)=n_0e^{-x/l_{\mathrm{diff}}}$

An applied non-homogeneous vertical electric field can alter the diffusion length (as demonstrated experimentally), which can be modelled as a change in the effective diffusion length.

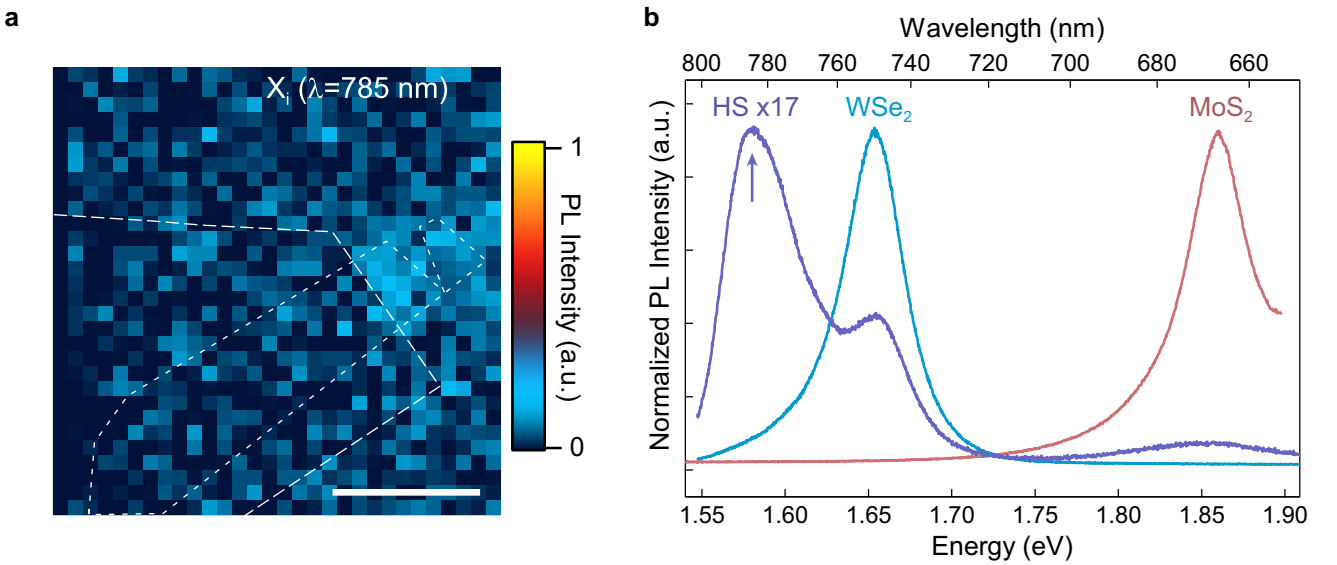
Numerical simulation of the exciton-energy profile. We first calculate the electric-field distribution in our system using the COMSOL Multiphysics simulation software. All calculations were performed considering the dimensions of the device as follows: the top graphene gates are 1.1 μm wide and spaced 0.8 μm apart. The heterostructure is encapsulated between two h-BN crystals (10 nm thick on the top and 20 nm on the bottom), and the substrate is heavily doped Si with 270 nm of SiO_2 on top (see Extended Data Fig. 15a). Extended Data Fig. 15b shows an example of the electrical field in the system in the confinement configuration, with −10 V applied to the central gate and the side gates grounded. Interlayer excitons have a built-in out-of-plane dipole moment directed upwards, with an absolute value of $p_z=ed=e\times7.5\times10^{-10}$ m, where e is the elementary charge and $d=7.5$ Å is the layer separation in our heterostructure. They thus experience an energy shift of $\delta E=-p_zE_z$ in the presence of a vertical electric field E_z. The resulting force applied on the exciton in the longitudinal direction is proportional to the first derivative of the vertical electric field E_z with respect to the channel x axis:

$$F_x=-\frac{\partial(\delta E)}{\partial x}=ed\frac{\partial E_z}{\partial x}$$

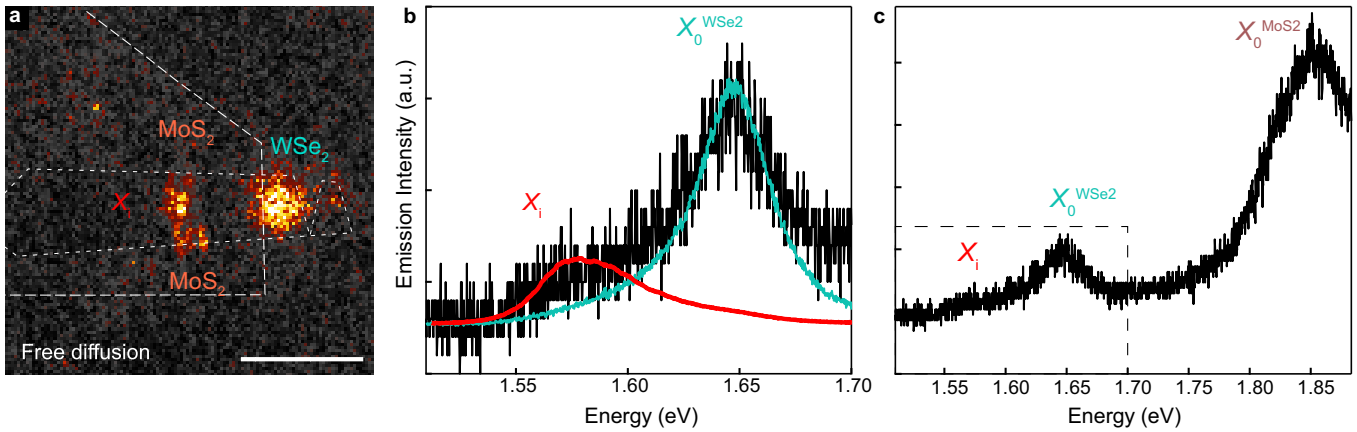
Example profiles of the confinement-well configuration are shown in Extended Data Fig. 15c.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

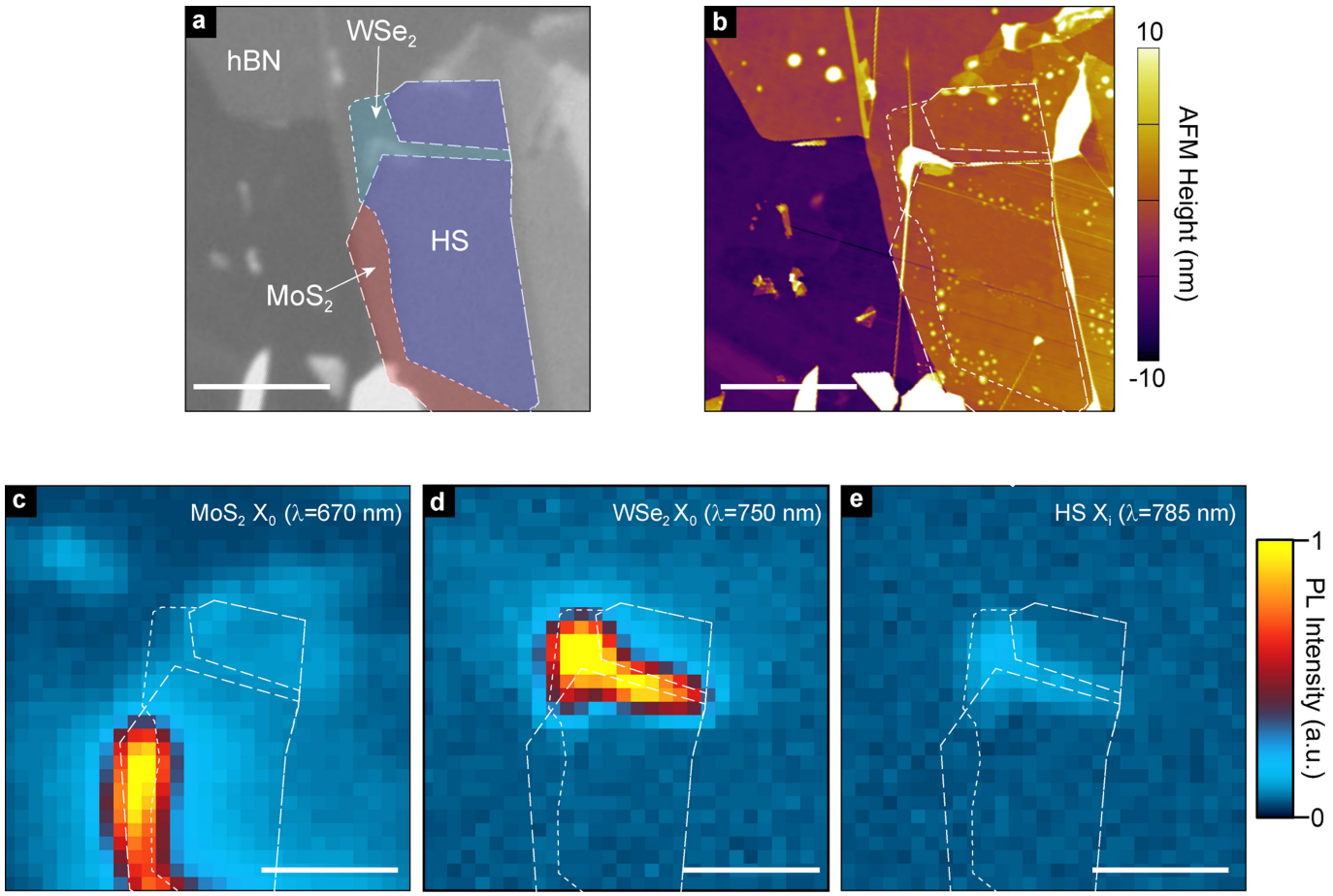
30. Mayorov, A. S. et al. Micrometer-scale ballistic transport in encapsulated graphene at room temperature. *Nano Lett.* **11**, 2396–2399 (2011).
31. Zhu, H. et al. Interfacial charge transfer circumventing momentum mismatch at two-dimensional van der Waals heterojunctions. *Nano Lett.* **17**, 3591–3598 (2017).
32. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).



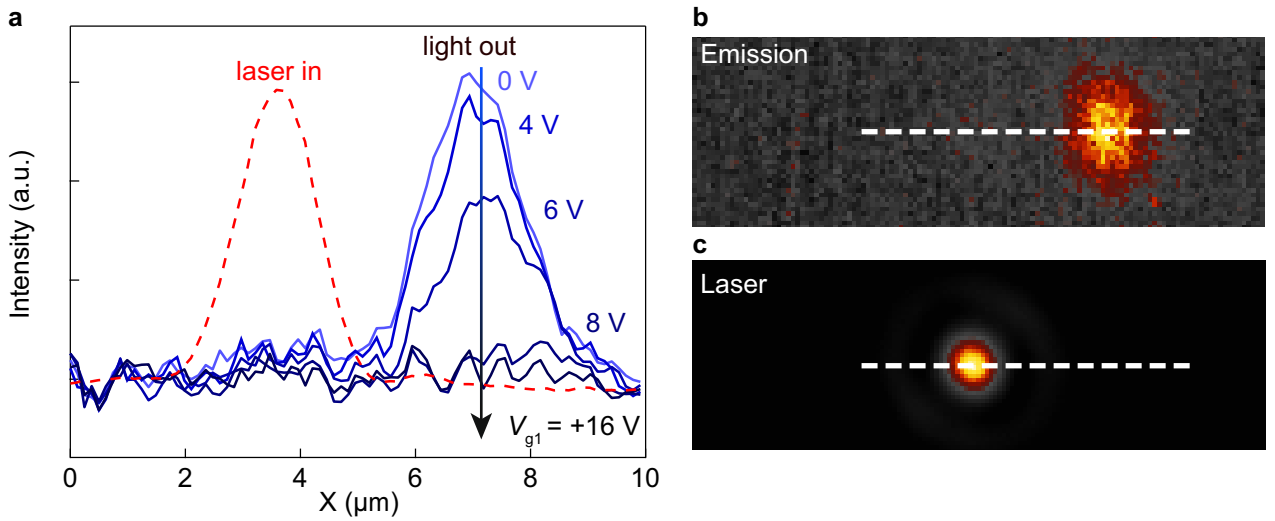
Extended Data Fig. 1 | Interlayer excitons in the WSe_2–MoS_2 van der Waals heterostructure. a, Spatial map of photoluminescence at 785 nm corresponding to the heterostructure interlayer photoluminescence emission maximum, as shown in the photoluminescence (PL) spectra in **b**. An efficient interlayer charge-transfer process in the heterostructure encapsulated in h-BN results in further quenching of photoluminescence emission from the heterostructure. Scale bar, 5 μm. **b**, Photoluminescence spectra from the structure fabricated on SiO_2.



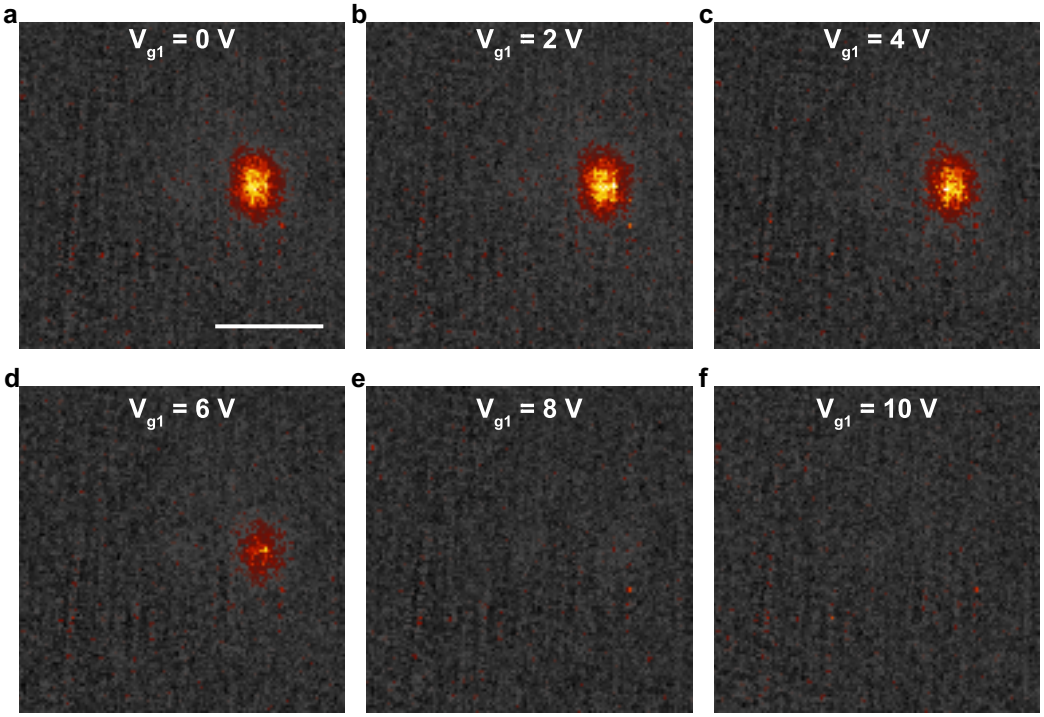
Extended Data Fig. 2 | Spectra of excitonic device emission. **a**, Distribution of photoluminescence emission intensity from the device, in the absence of an electric field. White dashed lines represent edges of constituent crystals. Scale bar, 5 μm. **b**, Detailed spectrum of the emission pattern, showing the interlayer exciton peak (X_i) and WSe_2 intralayer emission. We note that the low-energy peak (X_i) cannot be related to localized excitons in WSe_2, because they are observed only at cryogenic temperatures. **c**, Full spectrum of the emission shown in **a**, also showing the emission from MoS_2 ($X_0^{MoS_2}$), which is blocked by the filter in the CCD image. The black dashed box indicates the range of energies shown in **b**.



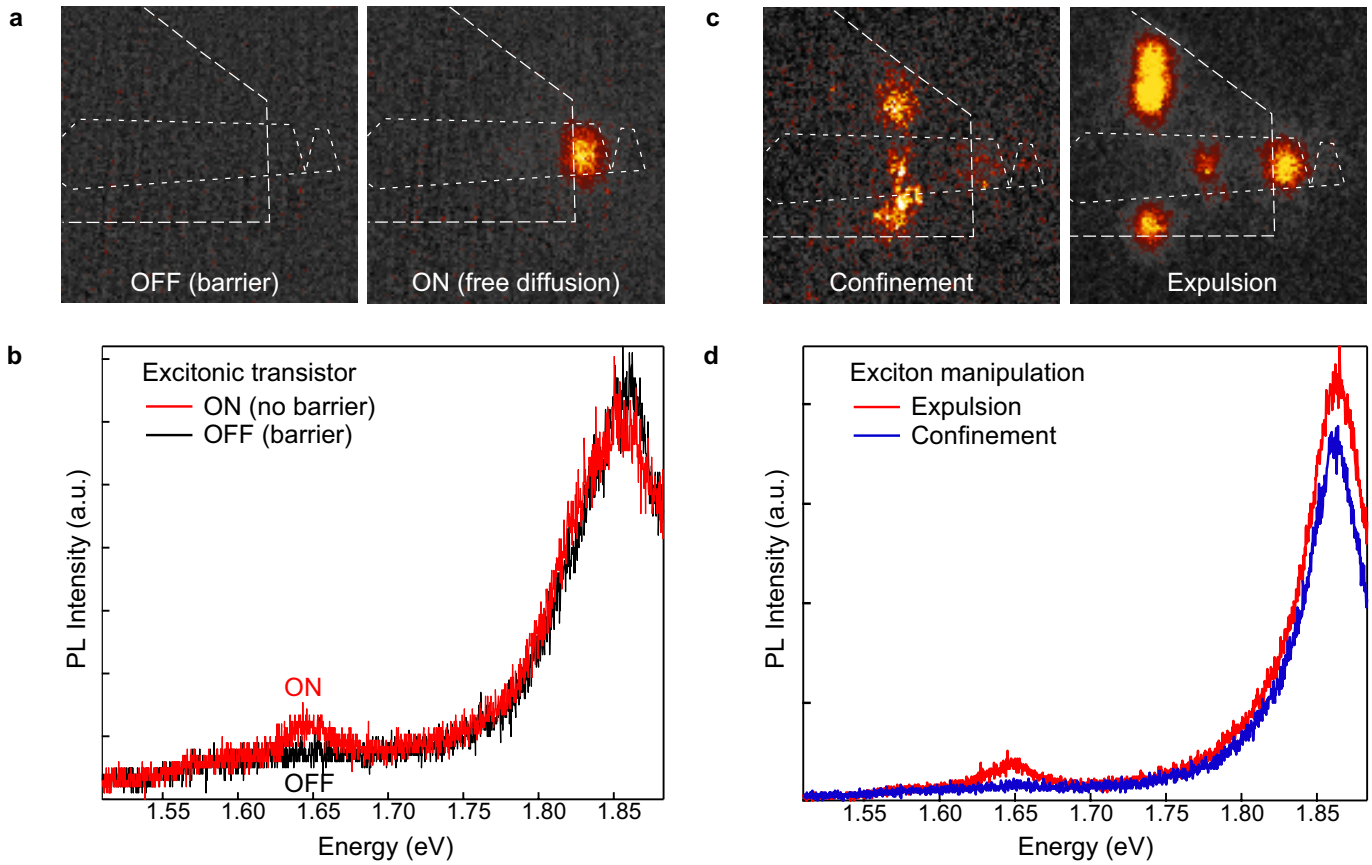
Extended Data Fig. 3 | Characterization of an additional WSe_2–MoS_2 heterostructure. a, False-colour optical image of the fabricated stack. **b**, Atomic force microscopy (AFM) height-profile image of the heterostructure. **c**–**e**, Spatial maps of photoluminescence intensity at emission wavelengths (λ) of 670 nm (**c**), 750 nm (**d**) and 785 nm (**e**), corresponding to MoS_2 intralayer (X_0), WSe_2 intralayer (X_0) and heterostructure interlayer (X_i) excitonic resonances. Photoluminescence is quenched in the heterostructure area owing to efficient charge transfer. White dashed lines represent edges of constituent crystals. Scale bars, 5 μm.



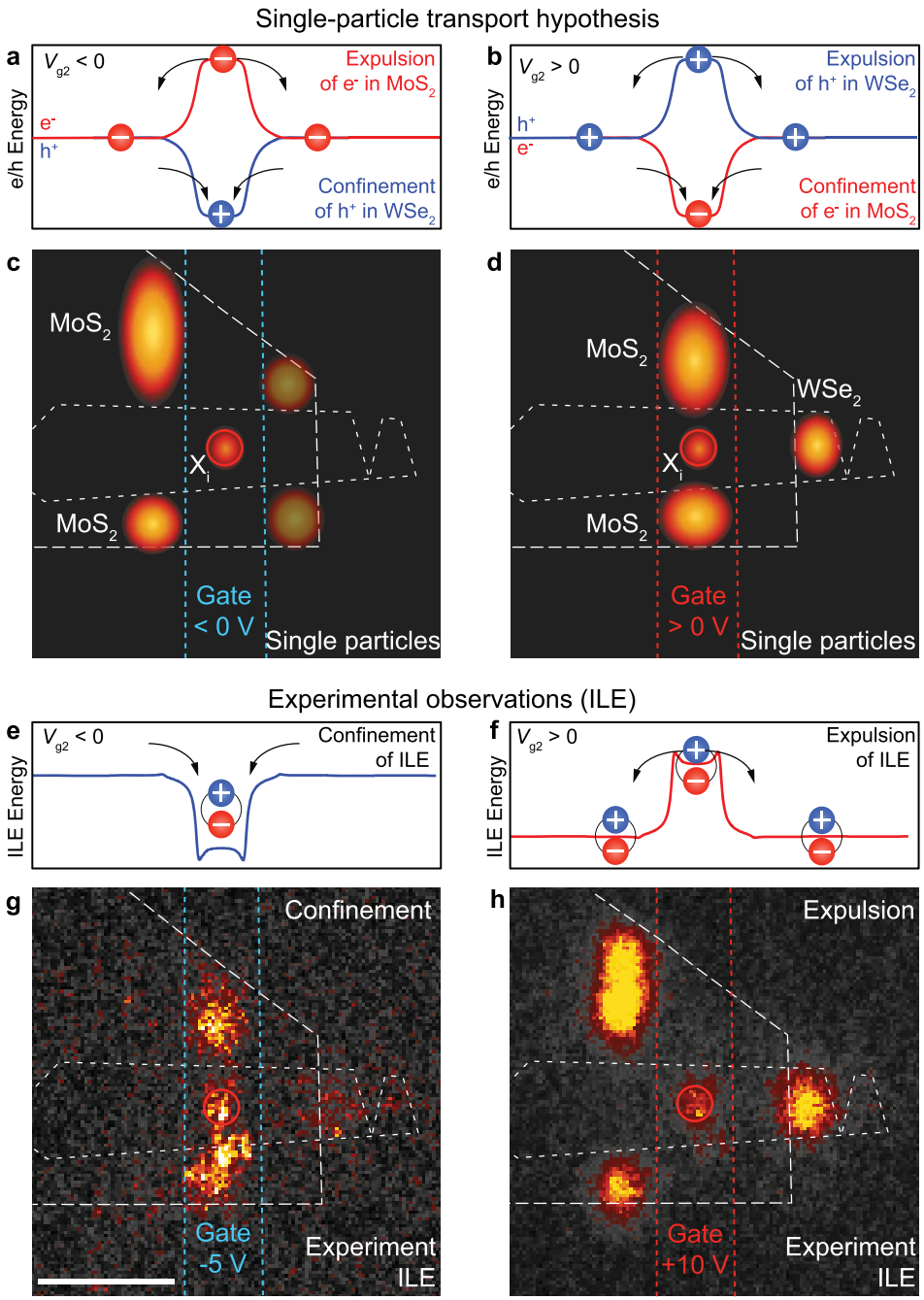
Extended Data Fig. 4 | Excitonic transistor input and output. a, Cross-sectional profile of the device emission intensity along the white dashed lines in **b** and **c**, obtained for different gate voltages V_{g1} from 0 V (light blue) to 16 V (black) with intermediate values of 4 V, 6 V and 8 V. The red line represents the intensity profile of the laser spot. **b**, **c**, CCD images of the exciton emission in the ON state (**b**) and of the focused laser spot (**c**). The lengths of the dashed lines indicate 10 μm.



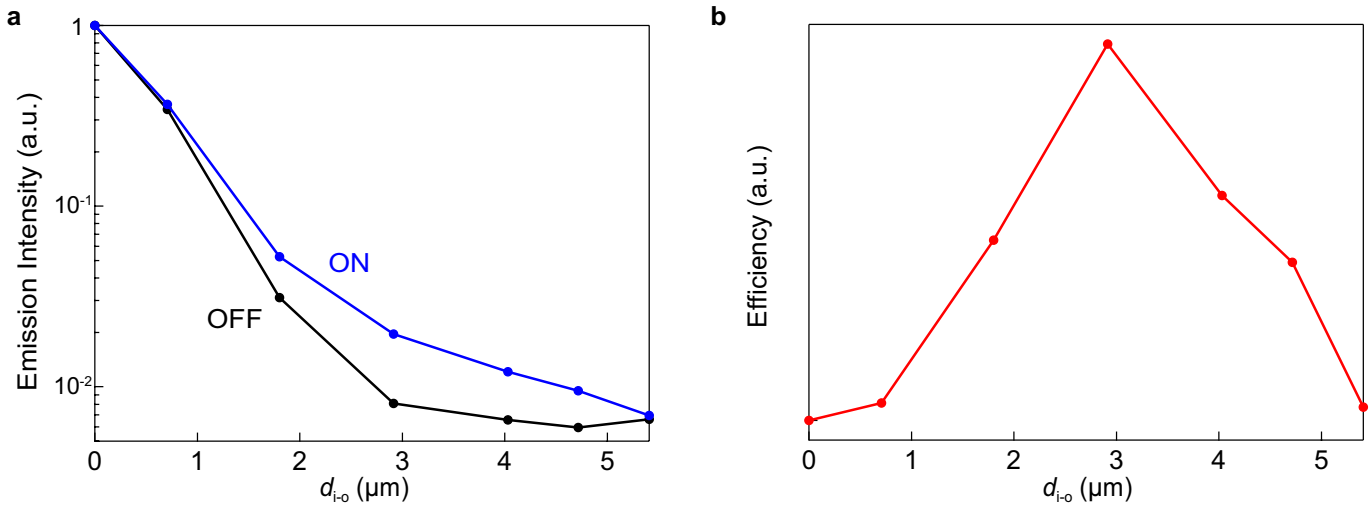
Extended Data Fig. 5 | Switching of the excitonic transistor. a–f, CCD images of exciton emission from the device, obtained for different gate voltages V_{g1} from 0 to 10 V in steps of 2 V.



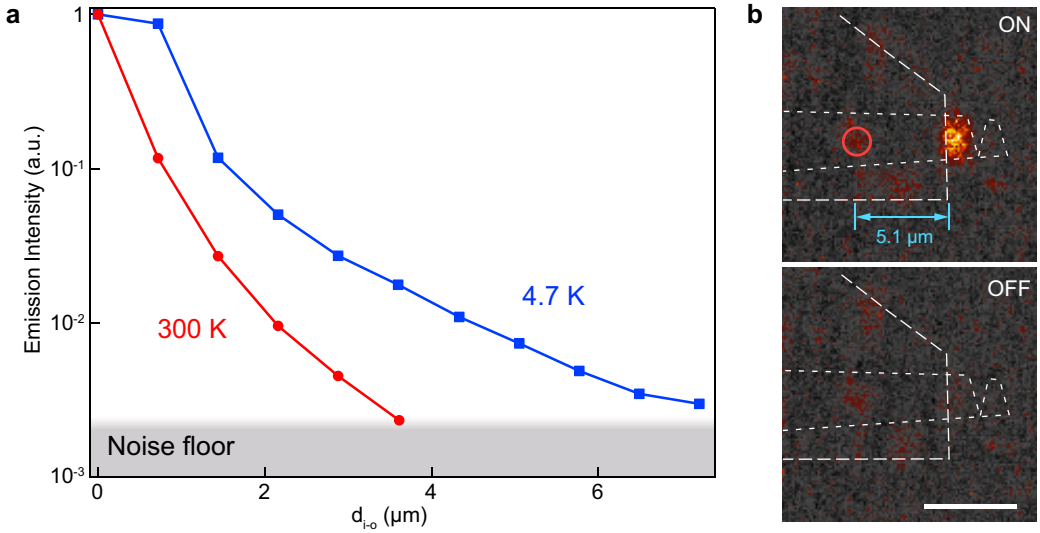
Extended Data Fig. 6 | Spectrum of light emitted from the device in different states. a, **b**, Intensity distribution of light emission from the excitonic transistor in the OFF and ON states (left and right, respectively; **a**) and the corresponding spectra collected from the entire device (black and red, respectively; **b**). **c**, **d**, Intensity distribution of light emission from the excitonic device in the confinement and expulsion configurations (left and right, respectively; **c**) and the corresponding spectra collected from the entire device (blue and red, respectively; **d**).



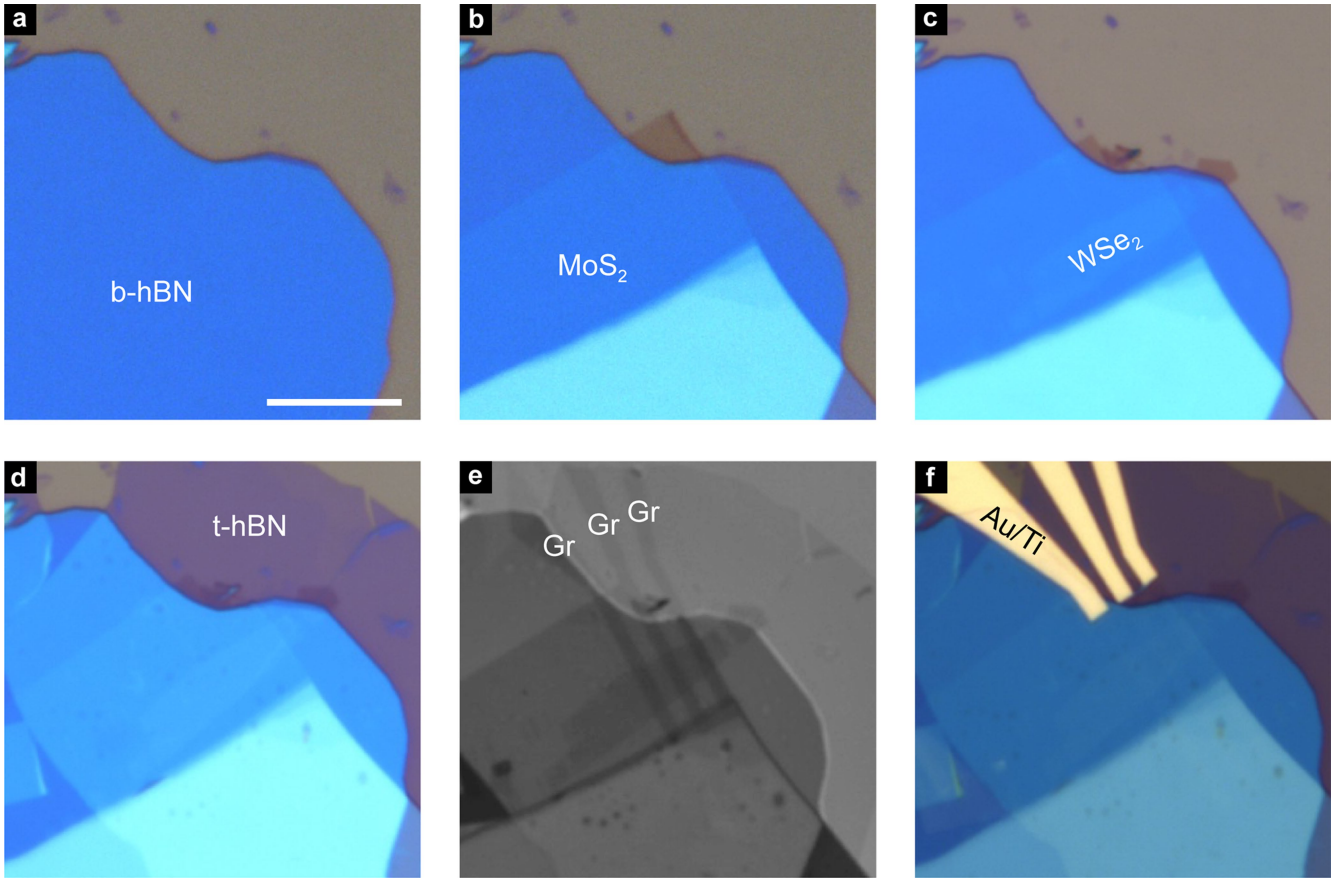
Extended Data Fig. 7 | Schematic depiction of the control over the light emission. a, **b**, Energy profiles for electrons (red) and holes (blue) in the device when applying negative (**a**) or positive (**b**) voltage on the central gate (V_{g2}). **c**, **d**, Corresponding expected emission images under the single-particle assumption. **e**, **f**, Energy profiles of interlayer excitons (ILE) in the presence of an external electric field, under the same conditions as in **a** and **b**. **g**, **h**, Corresponding experimental results. Scale bar, 5 μm. The figures in **a**–**d** are schematics based on the hypothesis that after the fast interlayer charge transfer, photo-excited carriers move independently rather than being bound in interlayer excitons. The diffusion of single electrons and holes is then subject to the type-II band alignment between MoS_2 and WSe_2, which restricts the motion of electrons to MoS_2 and holes to WSe_2. This charge separation is very efficient, as indicated by the strong suppression of intralayer emission from the heterostructure (Fig. 1e, f). Once the separation occurs, it is not very likely that the charges can hop between the layers: the band difference between MoS_2 and WSe_2 is more than 200 meV, so thermal excitation of 25 meV will not be enough for electrons to jump back to WSe_2 and holes to jump back to MoS_2. Another thing to consider is the local electrostatic potential defined by the gate. The application of $V_{g2} < 0$ creates a confining energy profile for single holes and a repulsive one for single electrons, as in **a** and **c**. Holes would then be confined in the WSe_2 area under the gate while electrons would be pushed out to MoS_2 areas next to the gate, where they would recombine with charges already present in the monolayer area, resulting in photoluminescence from single-layer areas of MoS_2 next to the gate (provided that there are enough holes in MoS_2 to start with). We would then expect to obtain the emission pattern shown in **c**, assuming the presence of native holes in MoS_2. In their absence, we would expect to see only one emission spot, coinciding with the excitation laser spot. Along the same lines, applying a positive gate voltage to the middle gate ($V_{g2} > 0$) would result in a repulsive potential for holes in WSe_2 and an attractive one for electrons in MoS_2. Recombination would then occur for electrons in MoS_2 in regions under the gate and for holes in WSe_2 in regions outside the gate, as shown in **d**. This is in contradiction with the experimental observations in **e**–**h**. In the case of interlayer exciton transport, we instead have only a single energy profile (**e**, **g**), and the application of a positive voltage on the middle gate results in the expulsion of interlayer excitons from the injection region (**f**, **h**).



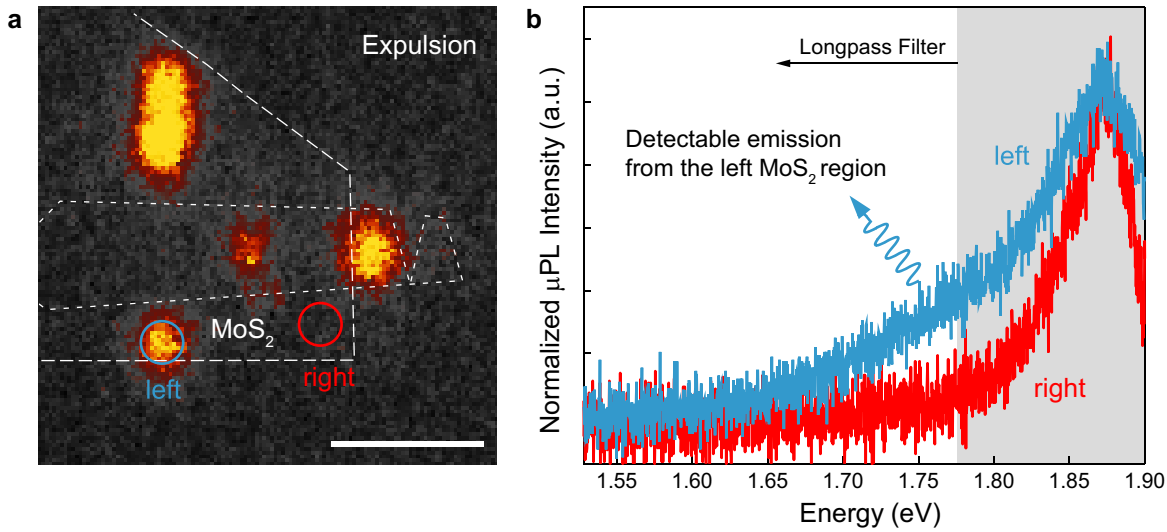
Extended Data Fig. 8 | Excitonic transistor characterization for different positions of the excitation laser spot. a, Normalized emission intensity (transistor output) as a function of the distance between optical injection and the emission point d_{i-o}, which is the same as in Fig. 3c, shown for the ON (blue, $V_{g1} = 0$ V) and OFF (black, $V_{g1} = 16$ V) states. **b**, Transistor efficiency calculated as the ratio between output emission in the ON and OFF states for different input–output separation distances d_{i-o}. Efficiency reaches a maximum when the laser spot is moved completely beyond the gate, so that the energy barrier stays between the input and the output and thus effectively modulates exciton diffusion.



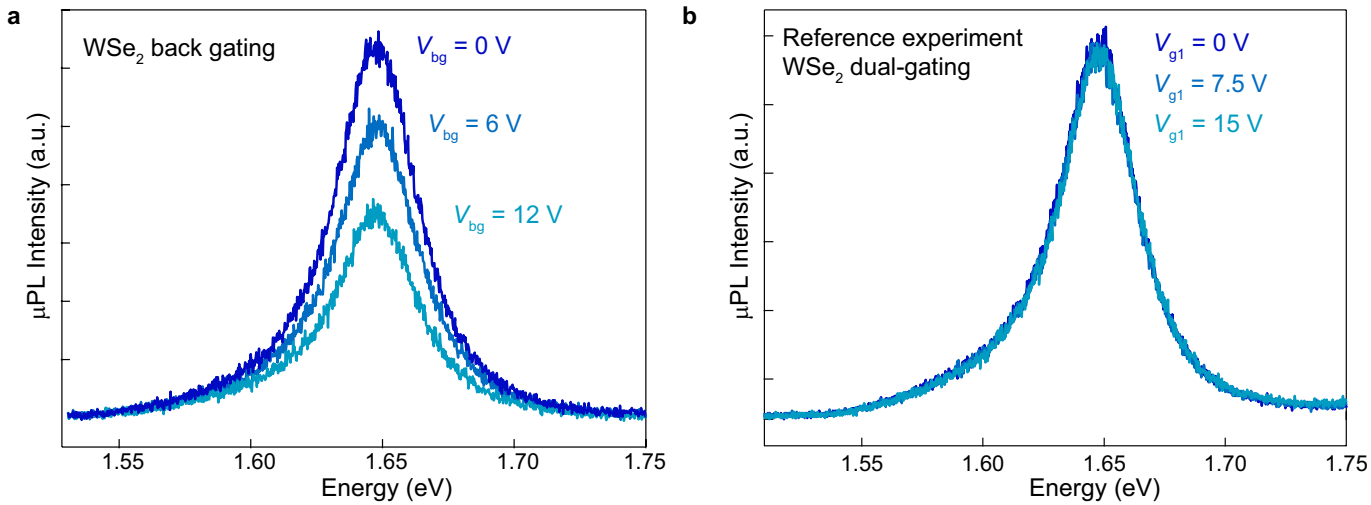
Extended Data Fig. 9 | Characterization of the device at low temperatures. a, Normalized output intensity as a function of the distance between optical injection and emission points, obtained at room temperature (red, 300 K) and 4.7 K (blue). No electric field is applied. **b**, Emission images of the device in the ON (top) and OFF (down) states when measured at 4.7 K, with input–output separations as long as $d_{i-o} = 5.1\,\mu m$. Such long-distance transistor switching was not observed at room temperature for this sample.



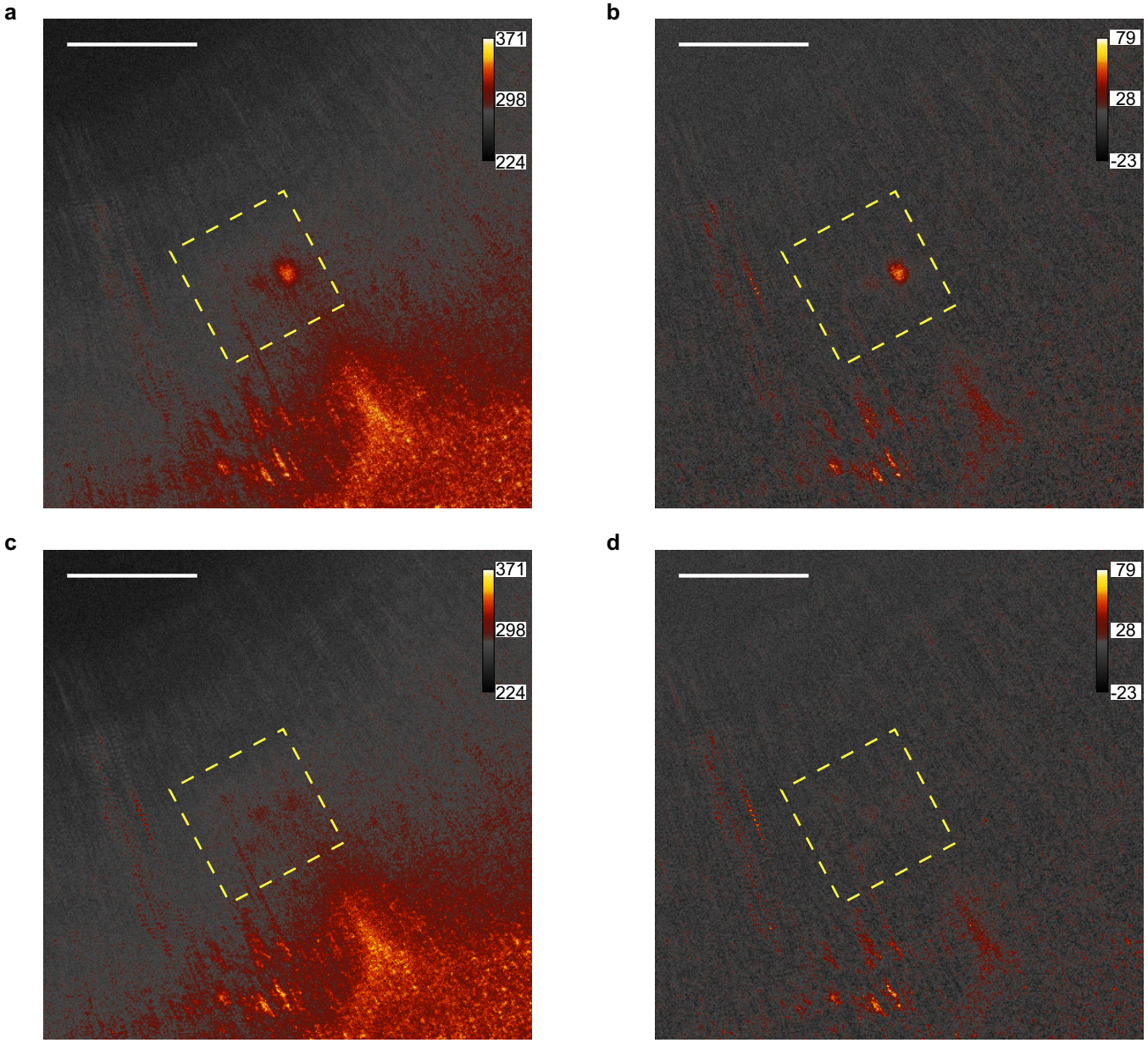
Extended Data Fig. 10 | Heterostructure fabrication. Optical images taken during different fabrication steps: **a**, exfoliation of the bottom h-BN (b-hBN); **b**, transfer of a monolayer MoS_2 flake; **c**, transfer of a monolayer WSe_2 flake; **d**, encapsulation with top h-BN (t-hBN); **e**, transfer of pre-patterned few-layer graphene stripes (Gr); and **f**, metallization of Au/Ti contacts. The image in **e** is shown in black and white for better visibility of the final structure. Scale bar, 10 μm (applies to all images).



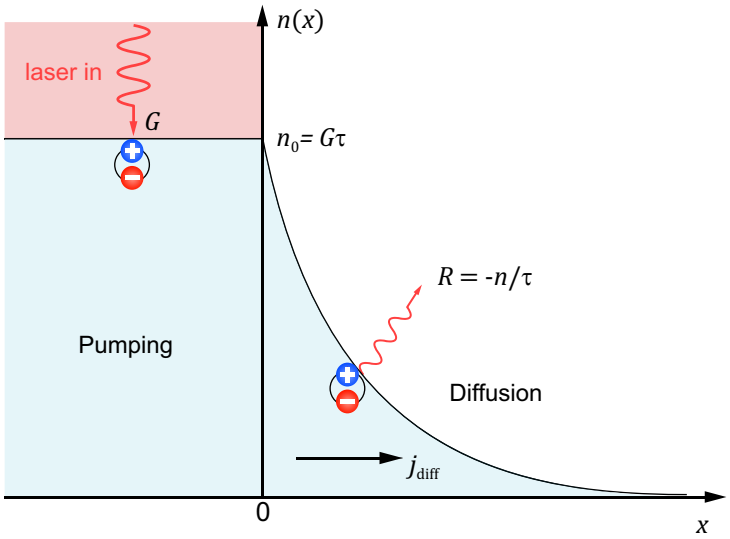
Extended Data Fig. 11 | Variation in photoluminescence emission from MoS_2 due to the inhomogeneity of the substrate. a, Image of photoluminescence emission coming from the device in the repulsive configuration shown in Fig. 4f. **b**, Micro-photoluminescence (μPL) spectra from the areas marked by red and blue circles in **a**, showing different peak widths as a result of local inhomogeneity in the heterostructure. The grey shaded area is the part of the spectrum cut by the 700-nm long-pass filter. As can be clearly seen in the image, areas where MoS_2 photoluminescence shows a low-energy tail due to broadening become visible to the CCD (left side of the device), whereas the other areas appear dark (right side).



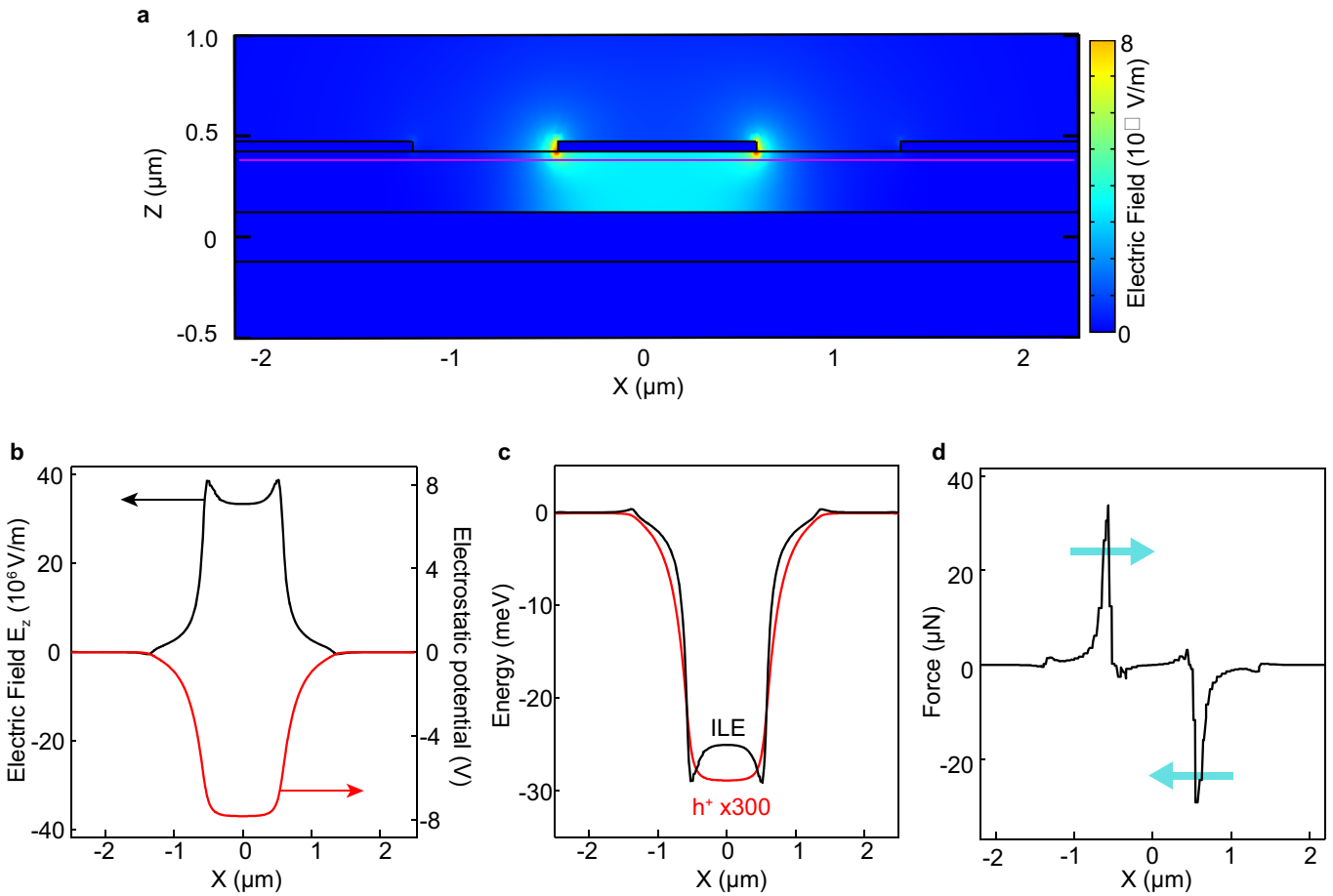
Extended Data Fig. 12 | Reference experiments. a, Photoluminescence spectra from monolayer WSe_2 at different back-gate voltages. Substantial modulation of the emission intensity is observed. **b**, Photoluminescence spectra from monolayer WSe_2 when using top and back gates in the dual-gated configuration, for the voltage range used in the experiment presented in Fig. 2. No appreciable intensity modulation is observed. Both measurements are performed on the same WSe_2 flake with the same continuous-wave excitation at 647 nm and 200 μW of incident power.



Extended Data Fig. 13 | Image post-processing. a, Original CCD image of the exciton emission for the configuration shown in Fig. 3a. The yellow square highlights the area of interest, shown in Fig. 3c. **b**, The same image after background subtraction. **c**, Original CCD image of the exciton emission for the configuration shown in Fig. 3b. **d**, The same image after background subtraction. Scale bars, 15 μm.



Extended Data Fig. 14 | Modelling of exciton diffusion. Schematic of exciton generation in the pumping area ($x < 0$) and diffusion outside ($x > 0$), represented by the exciton concentration $n(x)$. Constant pumping by the laser (red area) is determined by the generation rate G. This rate together with the recombination rate R establish the exciton concentration n_0 in the pumping region. The concentration gradient outside the pumping area generates an exciton flux j_{diff} that drives diffusion and leads to the exponential decay of exciton concentration along the x axis.



Extended Data Fig. 15 | Numerical simulation of the interlayer exciton in an external field. a, Two-dimensional cross-sectional map of the electric-field amplitude calculated for the device in the exciton confinement configuration, with −10 V applied to the central gate and the side gates grounded. **b**, Corresponding amplitude of the vertical electric field (black, left axis) and the electrostatic potential (red, right axis) along the heterobilayer. **c**, Energy shift experienced by an interlayer exciton (black, ILE) and a single hole (red, multiplied by 300) along the same cross-section. **d**, Projection along the *x* axis of the confinement force experienced by the interlayer exciton owing to the presence of the electric field. Blue arrows show the direction of the force.

LETTER

https://doi.org/10.1038/s41586-018-0397-3

Cryo–STEM mapping of solid–liquid interfaces and dendrites in lithium–metal batteries

Michael J. Zachman[1], Zhengyuan Tu[2], Snehashis Choudhury[3], Lynden A. Archer[2,3] & Lena F. Kourkoutis[1,4]*

Solid–liquid interfaces are important in a range of chemical, physical and biological processes[1–4], but are often not fully understood owing to the lack of high-resolution characterization methods that are compatible with both solid and liquid components[5]. For example, the related processes of dendritic deposition of lithium metal and the formation of solid–electrolyte interphase layers[6,7] are known to be key determinants of battery safety and performance in high-energy-density lithium-metal batteries. But exactly what is involved in these two processes, which occur at a solid–liquid interface, has long been debated[8–11] because of the challenges of observing such interfaces directly. Here we adapt a technique that has enabled cryo-transmission electron microscopy (cryo-TEM) of hydrated specimens in biology—immobilization of liquids by rapid freezing, that is, vitrification[12]. By vitrifying the liquid electrolyte we preserve it and the structures at solid–liquid interfaces in lithium-metal batteries in their native state, and thus enable structural and chemical mapping of these interfaces by cryo-scanning transmission electron microscopy (cryo-STEM). We identify two dendrite types coexisting on the lithium anode, each with distinct structure and composition. One family of dendrites has an extended solid–electrolyte interphase layer, whereas the other unexpectedly consists of lithium hydride instead of lithium metal and may contribute disproportionately to loss of battery capacity. The insights into the formation of lithium dendrites that our work provides demonstrate the potential of cryogenic electron microscopy for probing nanoscale processes at intact solid–liquid interfaces in functional devices such as rechargeable batteries.

Accurate high-resolution characterization of electrode–electrolyte interfaces is challenging owing to the volatility of commonly used liquid electrolytes, the high chemical reactivity of metal anodes such as lithium and the fact that the region of interest is an interface between two condensed phases of matter. To address this, the liquid is typically removed and the electrode of interest washed and dried before being characterized using traditional methods, which alters the structure and chemistry of the solid–liquid interface[13]. Here, we use cryogenic techniques originally designed for preserving hydrated biological specimens[12] coupled with cryo-focused ion beam (cryo-FIB) and analytical cryo-scanning transmission electron microscopy (cryo-STEM) techniques to access the intact structure and chemistry of dendrites and their interphases in lithium-metal batteries down to the nanoscale.

Figure 1a shows a schematic of the symmetric lithium-metal coin cells used for these experiments (for additional details, see Methods). To preserve the electrolyte on the electrode surface, the cells were opened and the electrode immediately plunge-frozen in a cryogen. To explore the morphology of the anode surface rapidly, we used cryo-FIB to mill a series of cross-sections through structures large enough to be localized by raised regions in the frozen electrolyte (Fig. 1b). We imaged each successive cross-section (Fig. 1c, d), revealing two distinct deposit morphologies, which we refer to as type I and type II dendrites. Type I dendrites are roughly 5 μm across with low curvature, whereas type II dendrites are generally hundreds of nanometres thick and tortuous. We did not observe any spatial correlations between dendrite types, nor cases where one dendrite type clearly formed on the other. To gain insight into their three-dimensional (3D) morphology, we reconstructed the 3D structure of the dendrites from the cross-sectional images (Fig. 1e), as has been demonstrated previously for biological samples[14]. The electrode contact areas for the individual structures can thereby be compared directly, revealing that the widths of type II dendrite contact areas are more than an order of magnitude smaller than those of type I dendrites. This suggests that type II dendrites may become disconnected from the electrode more easily during battery cycling and, in combination with their approximately equal numbers (Fig. 1f) and volumes, contribute disproportionately to capacity fade owing to electrochemically disconnected ('dead') lithium.

Although cryo-FIB techniques alone provide valuable morphological information, we used cryo-STEM and electron-energy-loss spectroscopy (EELS) to obtain high-resolution structural and chemical information about the dendrites and their associated solid–electrolyte interphase (SEI) layers. Electron-transparent cross-sectional lamellae were extracted from plunge-frozen anode–electrolyte interfaces using cryo-FIB lift-out[15] (Fig. 2a, b, Extended Data Fig. 1). High-angle annular dark-field (HAADF) cryo-STEM imaging immediately revealed an extended SEI layer on the type I dendrite, approximately 300–500 nm thick, which was not present on the type II dendrite (Fig. 2c, d). The SEI layer on lithium-metal battery anodes is generally thought to be tens of nanometres thick[7]. Our results suggest that a soft, extended portion of the SEI is removed by the typical washing and drying sample-preparation steps. The remaining SEI material observed by these techniques would then be a thin, compact layer. This is important, in part because it means substantially more lithium is irreversibly lost to the SEI layer than previously thought.

Spectroscopic mapping by EELS shows that the extended type I SEI has an increased concentration of oxygen and lithium compared to the electrolyte and contains essentially no fluorine (Fig. 2e). Although no extended type II SEI is present, there is a thin carbon-free, lithium- and oxygen-rich layer on the type II dendrite surface, approximately 20 nm thick (Fig. 2f). Roughly spherical structures up to micrometres in size, containing carbon, oxygen, lithium and an elevated level of fluorine (Fig. 2e), were frequently observed near both dendrite types (shown near the type I dendrite in Fig. 2a, c) but rarely elsewhere in the sample. Individual elemental maps, including nitrogen, are shown in Extended Data Fig. 2.

By analysing the fine structure of the carbon K-edge using multivariate curve resolution, we observe distinct carbon-bonding environments in the electrolyte, SEI and fluorine-rich structure (Fig. 3a, b, Extended Data Fig. 2b). The increased intensity from C=O bonds in carbonates[16] in the SEI, along with the increased oxygen content and reduced concentration of C–H bonds[16], is consistent with evidence that the SEI consists largely of lithium ethylene dicarbonate in ethylene-carbonate-based electrolytes[9,17,18]. In addition, ethylene gas is produced during the formation of lithium ethylene dicarbonate from ethylene carbonate[9,17–19], which may explain the large bubbles observed in the

[1]School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA. [2]Department of Materials Science and Engineering, Cornell University, Ithaca, NY, USA. [3]Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA. [4]Kavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, NY, USA.
*e-mail: lena.f.kourkoutis@cornell.edu

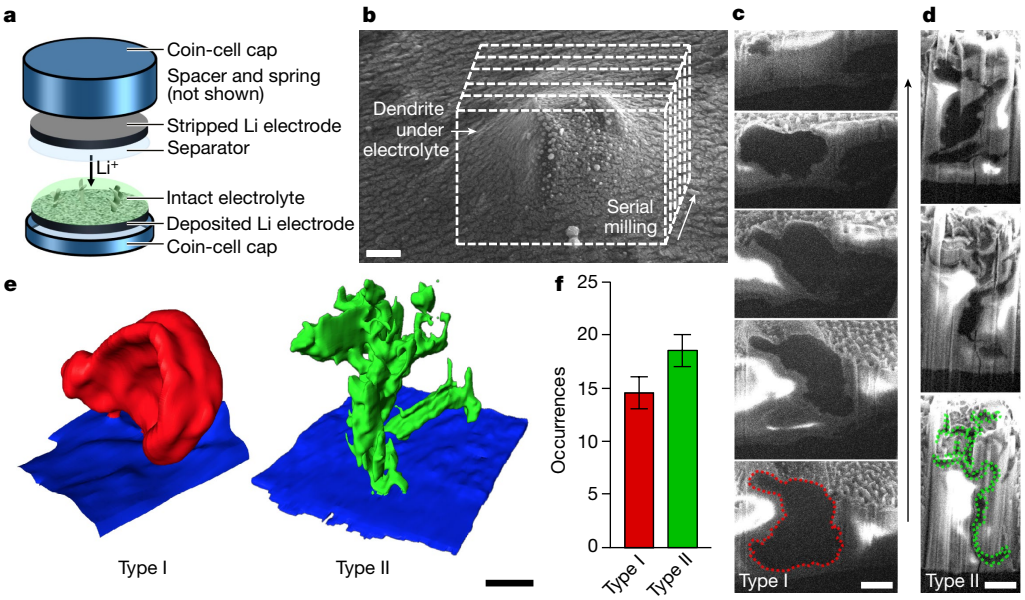


Fig. 1 | Characterization of dendrite morphologies by cryo-FIB. **a**, Coin-cell arrangement used. **b**, Raised regions in the electrolyte frozen on opened coin-cell electrodes reveal buried dendrite locations. The electrolyte surface was sputter-coated with a thin metal layer for increased conductivity. **c**, **d**, Two distinct dendrite morphologies, referred to as type I (**c**) and type II (**d**), were observed in serial cross-sectional images produced by cryo-FIB and cryo-scanning electron microscopy (cryo-SEM). **e**, Three-dimensional reconstructions of the dendrite structures highlight the morphological differences. **f**, Roughly equal numbers of the two morphologies were present across many coin cells. Error bars represent dendrites that were not unambiguously identified. Scale bars, 2 μm (**b**–**d**), 5 μm (**e**).

SEI. The correlation of fluorine and C = C bonds[16] in the fluorine-rich structure may indicate that ethylene is bound to LiF here, the possibility of which has been discussed previously[20]. Although no substantial changes in carbon bonding are observed at the type II dendrite surface (Fig. 3c, d), the fine structure of the oxygen K-edge of the thin type II SEI is consistent with lithium oxide and hydroxide monohydrate (Extended Data Fig. 3a).

EELS of the dendrite interiors show that the type I dendrite contains an appreciable quantity of oxygen, whereas the type II dendrite does not. The two dendrite types are therefore distinct in composition as well as morphology. The fine structures of the lithium and oxygen K-edge (Fig. 4a, b) reveal that the type I dendrite is composed primarily of lithium metal, partially oxidized. The lithium K-edge of the type II dendrite, however, unexpectedly corresponds to pure lithium hydride. Although hydrogen gas is known to be prevalent in cycled lithium batteries[21], only small amounts of LiH have been observed on freshly exposed lithium put in contact with organic electrolytes[22]. However, appreciable quantities have been observed on metal oxide conversion cathodes in lithium-ion batteries[13]. Although LiH is only metastable in electrolytes because it reacts rapidly with trace moisture and solvent molecules to form LiOH and Li_2O[22], these reactions may result in a thin passivating layer on the surface of large LiH structures that preserves the interior material, consistent with the type II dendrite.

Although it might be tempting to assume that the hydrogen required to produce the type II dendrites originates solely from reduction of water impurities in the electrolyte, it was shown recently that the decomposition of electrolyte solvent molecules can produce many times more hydrogen than water impurities provide[19]. By assuming a type II dendrite volume of about 300 μm^3, as determined by our cryo-FIB measurements, a quick calculation reveals that the maximum density of type II dendrites that could form in our cells as a result of hydrogen from water in the electrolyte (less than 10 p.p.m. H_2O) should be roughly an order of magnitude lower than what is actually observed. This suggests that the electrolyte solvent molecules may also be contributing hydrogen. Potential pathways to hydrogen production exist for any hydrogen-containing electrolyte, and therefore all lithium-metal batteries that use traditional organic electrolytes are likely to suffer from LiH dendrite formation and the associated capacity fade. However, the rate of hydrogen production from solvent molecules is a strong function of cell voltage[19], so full cell batteries that use higher-voltage cathodes would produce much larger quantities of hydrogen. This would exacerbate the problem of LiH dendrites in these cells, especially when 5-V-class high-voltage cathode materials that are designed to improve energy density are used.

We additionally examined the plasma resonances of the materials by simultaneously acquired low-loss EELS (Fig. 4c–e). The type I dendrite plasmon suggests that the lithium is only partially oxidized, because appreciably oxidized lithium forms additional resonances at 18 eV and 30 eV[23], which we did not observe. The approximately 13-eV shoulder on the peak corresponding to the type II dendrite also provides further evidence for the presence of LiH, because the hydrogen K-edge is found at 13.6 eV. Using the distinct low-loss spectra of the two dendrite materials, we map their spatial distribution within the dendrites (Fig. 4c, d). This mapping demonstrates that small LiH regions are also present on the surface of the type I dendrite and that a lithium particle is present on the tip of the type II dendrite. These results are summarized in Extended Data Table 1 and Extended Data Fig. 2d.

The lithium particle on the tip and the uniformity of the LiH within the type II dendrites, as well as their aspect ratio, suggest a root or tip growth mode. Grain boundaries have been shown to increase hydrogen diffusion through some metals[24], and hydrogen selectively penetrating into an electrode grain boundary or a lithium-particle–electrode interface could initiate LiH formation. The resulting volume expansion would lift the particle or electrode grain away from the electrode surface, leading to the observed lithium tip on the dendrite (Fig. 4d). Growth would then probably proceed mainly at the dendrite–electrode interface, owing to the poor electrical conductivity of LiH[25]. Although the formation of LiH is reversible[13], the reverse reaction would also occur primarily at the base. Furthermore, LiH is much more brittle than lithium metal[26]. These facts, combined with the low electrode contact area of the type II dendrites, suggest that the type II dendrites become disconnected from the electrode more easily during cycling than do type I dendrites. Because the total volume and number of dendrites are comparable for both types, type II dendrites may, therefore, contribute disproportionately to capacity fade by orphaned or disconnected lithium. The thickness of the type I SEI layer also means that it contributes

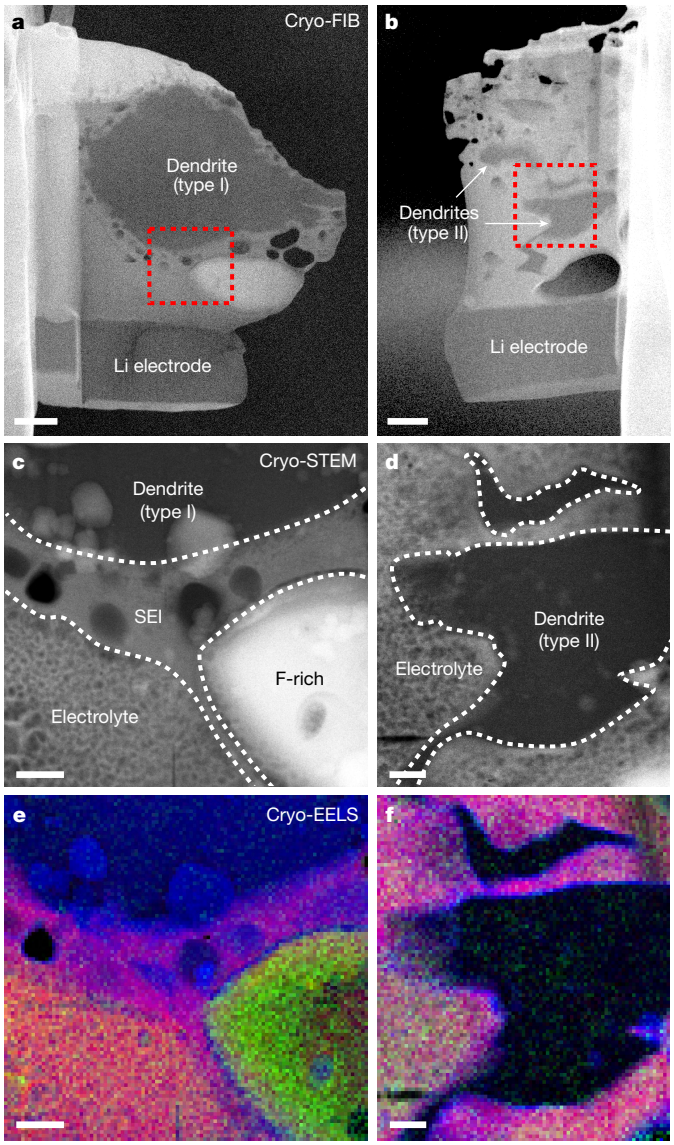


Fig. 2 | Structure and elemental composition of dendrites and their interphase layers in electron-transparent lamellae. a, **b**, Electron-transparent cryo-FIB lift-out lamellae of type I (**a**) and type II (**b**) dendrites. **c**, **d**, HAADF cryo-STEM imaging reveals an extended SEI layer on the type I dendrite (**c**), but not on the type II dendrite (**d**). **e**, **f**, EELS elemental mapping shows that both SEIs are oxygen-rich, but that the type II SEI contains no carbon (contrast has been adjusted for clarity; raw data are shown in Extended Data Fig. 2). The type I dendrite has an appreciable oxygen content (**e**), whereas the type II dendrite does not (**f**). Fluorine-rich structures were often observed near both dendrite types. Scale bars, 1 μm (**a**, **b**), 300 nm (**c**–**f**).

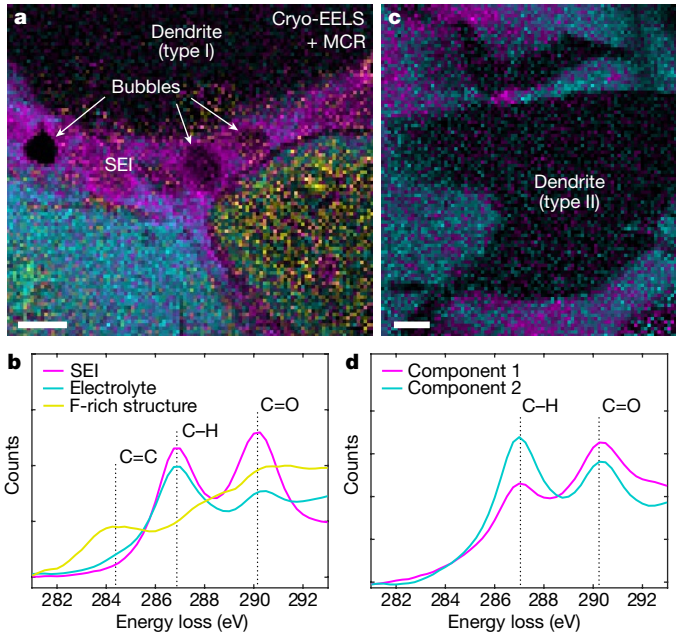


Fig. 3 | Analysis of the carbon-bonding environment near the dendrites. a, Spatial variation of the fine structure of the carbon K-edge near a type I dendrite, obtained from EELS, showing distinct carbon-bonding environments, as determined using multivariate curve resolution (MCR; contrast has been adjusted for clarity; raw data are shown in Extended Data Fig. 2). **b**, The increased carbonate C = O peak is consistent with a lithium ethylene dicarbonate SEI layer. Ethylene gas produced during SEI formation may explain the SEI bubbles seen in **a** and the C = C peak in the fluorine-rich structure. **c**, **d**, Two carbon components were present in the electrolyte around the type II dendrite, but were not localized along the dendrite surface. Scale bars, 300 nm.

more to the loss of lithium material than previously thought; we estimate, however, that an order of magnitude more lithium is contained within the type II dendrites than in the type I SEI layers. Minimizing the formation of LiH dendrites is therefore critical to improving the longevity of lithium-metal batteries.

Our results suggest that preventing the formation of type II dendrites may hinge on a careful choice of solvents and salts to eliminate hydrogen-containing species in the electrolyte and to form interphase layers that are better able to protect the anode. One way to achieve these goals would be to replace hydrogen in the solvent molecules with other elements to generate alternative species in the electrolyte. For example, we hypothesize that a properly chosen material would result in a hydrogen-deprived and fluorine-rich environment in which sacrificial, low-stability-window fluorinated components preferentially react at the electrodes relative to carbonate solvents, resulting in fluorine-rich species in the electrolyte. This would both starve the system of hydrogen and promote the formation of a beneficial LiF-rich barrier layer on the surface of the anode[27,28], stunting the growth of LiH dendrites and reducing capacity fade. Although recent studies that used high-concentration full-fluoride electrolytes are consistent with our predictions[29], maintaining the high fluorine-donating salt content of such electrolytes is currently impractical because of the high cost of the salt. There is therefore reason to explore alternative hydrogen-free electrolyte materials that are as effective using different salts at lower concentrations.

As a test of the above hypothesis, we performed cryo-FIB, cryo-STEM EELS and electrochemical experiments on cells prepared with a full-fluoride-type electrolyte using a lower salt concentration of a less expensive salt (2 M $LiPF_6$) than in previous reports; a fully fluorinated solvent, fluoroethylene carbonate, was used in these electrolyte compositions. The results are shown in Extended Data Fig. 4. Consistent with our hypothesis, we found that this fluorinated electrolyte suppresses the formation of LiH dendrites substantially and greatly alters the lithium deposition. Localized structures are still present, but rather than type I dendrites the deposits are much larger, with an inner structure consisting of many smaller 'blocks' separated by SEI layers. These blocks are composed of partially oxidized lithium, as with the type I dendrites, which was confirmed by cryo-STEM EELS of structures prepared by cryo-FIB lift-out. The electrochemical performance was correspondingly enhanced, with higher Coulombic efficiency and greatly reduced capacity fade, as shown in a lithium versus stainless steel set-up and in a full-cell battery that used a nickel manganese cobalt oxide cathode. Although this demonstrates the feasibility of dendrite suppression and improved battery performance by introducing fluorinated electrolytes, the concepts outlined above could potentially be further developed by integrating the alternative electrolyte species into cross-linkable structures, such as those reported recently[30]. This would result in a

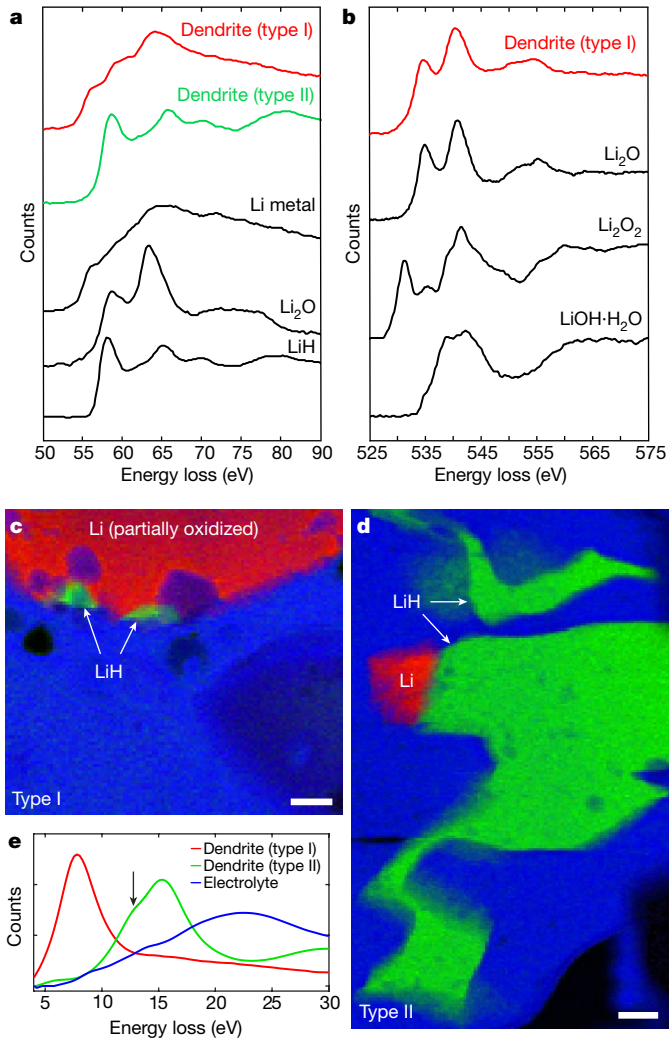


Fig. 4 | Determination and mapping of dendrite composition. **a**, **b**, Comparison of the fine structures of the lithium (**a**) and oxygen (**b**) K-edge of the dendrites (coloured lines) with reference materials (black) reveals that the type I dendrite was partially oxidized lithium metal, whereas the type II dendrite was uniform lithium hydride. **c**–**e**, Mapping of the lithium metal and hydride low-loss EELS spectra (**e**) reveals that the type I dendrite was only slightly oxidized and had small LiH regions at its surface (**c**), whereas the type II dendrite had a lithium particle tip (**d**) (contrast has been adjusted for clarity; raw data are shown in Extended Data Fig. 2). The arrow in **e** denotes the hydrogen K-edge that appears at about 13 eV in LiH. Scale bars, 300 nm.

hydrogen-deprived and halide-rich electrolyte environment that simultaneously form lithium-halide-rich and elastic interphases that are able to flex to accommodate changes in the volume of the lithium anode, thus providing further protection against both LiH and traditional lithium-metal dendrites. Cryo-FIB and analytical cryo-STEM techniques have thus provided us access to the nanoscale structure and composition of intact solid–electrolyte interfaces in lithium-metal batteries, revealing the existence of LiH dendrites and extended SEI layers on lithium metal dendrites, and allowed us to propose pathways to overcoming their detrimental effects.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0397-3.

Received: 17 January 2018; Accepted: 4 July 2018;
Published online 15 August 2018.

1. Wu, Y. & Yang, P. Direct observation of vapor–liquid–solid nanowire growth. *J. Am. Chem. Soc.* **123**, 3165–3166 (2001).
2. Weiner, S. & Addadi, L. Crystallization pathways in biomineralization. *Annu. Rev. Mater. Res.* **41**, 21–40 (2011).
3. Stamenkovic, V. R., Strmcnik, D., Lopes, P. P. & Markovic, N. M. Energy and fuels from electrochemical interfaces. *Nat. Mater.* **16**, 57–69 (2017).
4. Tarascon, J.-M. & Armand, M. Issues and challenges facing rechargeable lithium batteries. *Nature* **414**, 359–367 (2001).
5. Zaera, F. Probing liquid/solid interfaces at the molecular level. *Chem. Rev.* **112**, 2920–2986 (2012).
6. Tikekar, M. D., Choudhury, S., Tu, Z. & Archer, L. A. Design principles for electrolytes and interfaces for stable lithium-metal batteries. *Nat. Energy* **1**, 16114 (2016).
7. Cheng, X.-B. et al. A review of solid electrolyte interphases on lithium metal anode. *Adv. Sci.* **3**, 1500213 (2016).
8. Yamaki, J. et al. A consideration of the morphology of electrochemically deposited lithium in an organic electrolyte. *J. Power Sources* **74**, 219–227 (1998).
9. Aurbach, D. Review of selected electrode-solution interactions which determine the performance of Li and Li ion batteries. *J. Power Sources* **89**, 206–218 (2000).
10. Bai, P., Li, J., Brushett, F. R. & Bazant, M. Z. Transition of lithium growth mechanisms in liquid electrolytes. *Energy Environ. Sci.* **9**, 3221–3229 (2016).
11. Kushima, A. et al. Liquid cell transmission electron microscopy observation of lithium metal growth/dissolution: root growth, dead lithium and lithium flotsams. *Nano Energy* **32**, 271–279 (2017).
12. Dubochet, J. et al. Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).
13. Hu, Y.-Y. et al. Origin of additional capacities in metal oxide lithium-ion battery electrodes. *Nat. Mater.* **12**, 1130–1136 (2013).
14. Heymann, J. A. W. et al. Site-specific 3D imaging of cells and tissues with a dual beam microscope. *J. Struct. Biol.* **155**, 63–73 (2006).
15. Zachman, M. J., Asenath-Smith, E., Estroff, L. A. & Kourkoutis, L. F. Site-specific preparation of intact solid–liquid interfaces by label-free in situ localization and cryo-focused ion beam lift-out. *Microsc. Microanal.* **22**, 1338–1349 (2016).
16. Cody, G. D. et al. Quantitative organic and light-element analysis of comet 81P/Wild 2 particles using C-, N-, and O-mu-XANES. *Meteorit. Planet. Sci.* **43**, 353–365 (2008).
17. Yang, C. R., Wang, Y. Y. & Wan, C. C. Composition analysis of the passive film on the carbon electrode of a lithium-ion battery with an EC-based electrolyte. *J. Power Sources* **72**, 66–70 (1998).
18. Zhuang, G. V., Xu, K., Yang, H., Jow, T. R. & Ross, P. N. Lithium ethylene dicarbonate identified as the primary product of chemical and electrochemical reduction of EC in 1.2 M LiPF 6/EC:EMC electrolyte. *J. Phys. Chem. B* **109**, 17567–17573 (2005).
19. Metzger, M., Strehle, B., Solchenbach, S. & Gasteiger, H. A. Origin of H_2 evolution in LIBs: H_2O reduction vs. electrolyte oxidation. *J. Electrochem. Soc.* **163**, A798–A809 (2016).
20. Szczęśniak, M. M. & Ratajczak, H. Ab initio calculations on the lithium fluoride–ethylene complex. *J. Chem. Phys.* **67**, 5400–5401 (1977).
21. Onuki, M. et al. Identification of the source of evolved gas in Li-ion batteries using $^{\#2}$#1 -labeled solvents. *J. Electrochem. Soc.* **155**, A794–A797 (2008).
22. Aurbach, D. & Weissman, I. On the possibility of LiH formation on Li surfaces in wet electrolyte solutions. *Electrochem. Commun.* **1**, 324–331 (1999).
23. Liu, D.-R. & Williams, D. B. The electron-energy-loss spectrum of lithium metal. *Philos. Mag. B* **53**, L123–L128 (1986).
24. Oudriss, A. et al. The diffusion and trapping of hydrogen along the grain boundaries in polycrystalline nickel. *Scr. Mater.* **66**, 37–40 (2012).
25. Islam, A. K. M. A. Lighter alkali hydride and deuteride. *Phys. Status Solidi b* **180**, 9–57 (1993).
26. Settouti, N. & Aourag, H. Structural and mechanical properties of alkali hydrides investigated by the first-principles calculations and principal component analysis. *Solid State Sci.* **58**, 30–36 (2016).
27. Choudhury, S. & Archer, L. A. Lithium fluoride additives for stable cycling of lithium batteries at high current densities. *Adv. Electron. Mater.* **2**, 1500246 (2016).
28. Lu, Y., Tu, Z. & Archer, L. A. Stable lithium electrodeposition in salt-reinforced electrolytes. *Nat. Mater.* **13**, 961–969 (2014).
29. Suo, L. et al. Fluorine-donating electrolytes enable highly reversible 5-V-class Li metal batteries. *Proc. Natl Acad. Sci. USA* **115**, 1156–1161 (2018).
30. Zhao, Q. et al. Building organic/inorganic hybrid interphases for fast interfacial transport in rechargeable metal batteries. *Angew. Chem. Int. Ed.* **57**, 992–996 (2018).

Acknowledgements We thank D. Muller, H. Abruña and J. Noble for discussions. M.J.Z. and L.F.K. acknowledge support by the NSF (DMR-1654596) and the Packard Foundation. Z.T., S.C. and L.A.A. acknowledge support from the Department of Energy, Advanced Research Projects Agency - Energy (ARPA-E) through award number DE-AR0000750. This work made use of the Cornell Center for Materials Research (CCMR) Shared Facilities with funding from the NSF MRSEC programme (DMR-1719875). Additional support for the FIB/SEM cryo-stage and transfer system was provided by the Kavli Institute at Cornell and the Energy Materials Center at Cornell, DOE EFRC BES (DE-SC0001086). The FEI Titan Themis 300 was acquired through NSF MRI-1429155, with additional support from Cornell University, the Weill Institute and the Kavli Institute at Cornell. This work made use of electrochemical characterization facilities in the KAUST-CU Center for Energy and Sustainability, supported by

the King Abdullah University of Science and Technology (KAUST) through award number KUS-C1-018-02.

Author contributions M.J.Z. and L.F.K. conceived the project and wrote the manuscript. M.J.Z. prepared samples, performed the experiments and analysed the data. Z.T. constructed, cycled and disassembled the traditional electrolyte coin cell batteries. S.C. constructed, cycled, dissembled and measured the electrochemical properties of the full-fluoride and traditional electrolyte coin cell batteries displayed in Extended Data Figure 4b–f. Z.T. and L.A.A. provided assistance with interpretation of the data and made revisions to the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0397-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to L.F.K.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Instrumentation and experimental details. We used an FEI Strata 400S DualBeam focused ion beam/scanning electron microscope system (FIB/SEM) to characterize and prepare samples. It was fitted with a Quorum PP3010T cryo-SEM/FIB system, which included a liquid nitrogen cold stage and an anticontaminator in the main FIB chamber, a preparation chamber (for sputter coating) with a separate cold stage and an anticontaminator attached to the FIB and separated by a valve, a stand-alone workstation for freezing and loading samples onto the specimen shuttle, and a vacuum transfer device for transporting samples between the workstation and preparation chamber. In addition, we installed an Oxford OmniProbe Cryoshaft on our OmniProbe 200 nanomanipulator, which is thermally isolated from the room-temperature shaft by a ceramic section and cooled by a copper braid attached to the anticontaminator[15]. Preparation of lamellae by cryo-FIB lift-out was carried out using techniques described previously[15]. All milling was performed at an ion beam voltage of 30 kV. Trenches to form the initial lamella were generally milled with a beam current of a few nanoamps. Thinning of the lamella was first conducted with a beam current of hundreds of picoamps, decreasing with lamella thickness to a final thinning with tens of picoamps. After cryo-FIB lift-out preparation, cryo-STEM samples were transferred back into liquid N_2 in the workstation where they were loaded into cryogenic sample storage boxes and transferred to a large liquid N_2 storage dewar.

Cryo-STEM characterization of these samples was performed on an aberration-corrected FEI Titan Themis operated at 300 kV. The microscope was equipped with an X-FEG high-brightness gun and a high-resolution Gatan imaging filter (GIF Quantum 965) for EELS. Standard Gatan side-entry cryo-transfer holders (model 626 and model 915) enabled transfer of the samples into the microscope and maintained their temperature near −180 °C throughout the experiment. The samples were loaded into the holder under liquid N_2 to minimize ice contamination. During transfer into the vacuum of the microscope, the sample was enclosed by a cryo-shutter which minimized ice build-up. No EELS edges beyond those discussed in the main text were observed in the dendrites between the Li K-edge at 55 eV and about 730 eV, including nitrogen, which rules out reactions with air or liquid N_2 during specimen preparation and transfer (Extended Data Fig. 5). While throughput of the cryo-FIB lift-out/cryo-STEM workflow is continuing to be improved, it can approach that of room-temperature FIB and STEM techniques with proper optimization. Multiple lamellae were prepared for analysis by cryo-STEM for this project. Additional examples are shown in Extended Data Fig. 1i–k, including an uncycled electrode for reference, and two containing type II dendrites. The O K-edge reference spectra for Li_2O_2 and LiOH and the Li K-edge reference spectrum for LiH were taken on a 200-kV FEI F20 using the same cryo-transfer holders and loading techniques.

The probe current for EELS maps on the Titan was around 25 pA, confirmed by measurement on a direct electron detector with a high dynamic range[31], and pixel dwell times were 10–50 ms. The electron dose applied during acquisition of the spectroscopic maps shown in the main text was 5×10^1–5×10^2 e^- Å^{-2}. The small bubbles in the electrolyte appeared rapidly, beginning by the time the first image was taken, with a total dose below 10 e^- Å^{-2}. These bubbles were probably hydrogen liberated from the electrolyte solvent molecules[32], since C and O K-edge fine structures in carbonates are known to be stable under the beam up to a dose of about 750 e^- Å^{-2} at room temperature under a 200-keV electron beam[33]. The threshold damage for these materials under our cryogenic conditions using a 300-keV beam should be higher than this[32]. A series of maps of the electrode–electrolyte interface taken at various total doses is shown in Extended Data Fig. 6a, c, demonstrating the doses at which different types of damage occur. The damage mechanisms of these carbonate solvents are liberation of hydrogen at low doses, resulting in structural changes such as the bubbling observed. At doses of more than 10^3 e^- Å^{-2}, mass loss becomes important and the fine structure of the approximately 287-eV peak is affected. At 10^4 e^- Å^{-2} the mass loss is severe, producing holes in the sample and leaving behind mainly the carbonate portion of the solvent molecules. However, the fine structure associated with this part of the molecule survives high doses. These findings are summarized in Extended Data Table 2. On the basis of our damage analysis, we do not expect the fine structure to have been altered in the maps shown in the main text, although slight structural modifications were present, as expected (Extended Data Fig. 6d, e).

LiH spectra have been measured previously using various techniques, and are not all in agreement[34–36]. The moisture sensitivity of the material may explain the disagreement. While precautions were taken in previous work to avoid air exposure, such as cleaving the LiH in vacuum[34] or transferring it to the microscope in an argon bag[35], reactions with small amounts of contaminants could still occur at room temperature. Accurate spectra could be obtained either by measuring a bulk sample, as the passivating layer would enable the majority of the material to remain unreacted, or by maintaining the sample at cryogenic temperatures, since reactions with contaminants would be essentially eliminated. As expected, the spectra from samples measured in bulk or at cryogenic temperatures are in agreement with each other[34,36] and with our data. The difficulty of characterizing unaltered LiH may be one reason why LiH dendrites have not been observed before, and is an important example of how cryogenic techniques such as cryo-FIB lift-out and cryo-STEM enable accurate characterization of systems with reactive materials as well as of solid–liquid interfaces.

Acquisition of reference spectra. The reference spectra for the lithium metal and Li_2O samples were acquired on the Titan in similar conditions to those described above. The Li K-edge for lithium metal was recorded on an uncycled lithium electrode and the Li_2O spectrum was recorded on a lithium electrode oxidized in the microscope by warming to room temperature and exposing the electrode to the beam. Both samples were produced by cryo-FIB lift-out. The electron dose for the metal spectrum was 10^4 e^- Å^{-2} and no change in fine structure was recorded by doubling this dose. The oxide spectra were acquired with a total dose of around 10^2 e^- Å^{-2}. While we observed that other lithium–oxygen compounds converted to Li_2O under the beam, no change to the Li_2O fine structure was observed at high doses.

We acquired the lithium peroxide and hydroxide reference materials from Sigma Aldrich, crushed them into a fine powder using a mortar and pestle, and pressed a holey carbon TEM grid onto the powder to adhere some to the grid. The LiH was prepared in an argon-filled glove box owing to its air sensitivity, and removed in a sealed vial which was opened under liquid N_2, eliminating air exposure. The other stable samples were also immediately placed under liquid N_2 after preparation to minimize unnecessary air exposure. On the F20, we used a probe current of about 75 pA. A total dose of less than 10^4 e^- Å^{-2} was applied to the oxides during acquisition, which we found was a few times lower than the dose necessary to induce a substantial change in the O K-edge fine structure. The LiH Li K-edge spectrum was acquired with a total dose of the order of 10^3 e^- Å^{-2}, and no change in fine structure was observed under any dose, measured to greater than 10^4 e^- Å^{-2}. All of the threshold damages for materials relevant to this study and the corresponding damage mechanisms are shown in Extended Data Table 2, and examples of damage series profiles used to establish these values are shown in Extended Data Fig. 6a.

To align the energy axis between the O K-edge spectra acquired on the Titan and the F20, we used the Li_2O peak at about 535 eV. The Li K-edge spectra were close enough to the zero-loss peak that no shifting of the spectra was necessary. In addition, spectra acquired on the F20 were bandpass-filtered by 0.6 eV to reduce noise below the energy resolution of the microscope. This reduced high-frequency noise while preserving larger features accurately, with only a slight reduction in sharp peaks. An example is shown for the Li_2O_2 O K-edge in Extended Data Fig. 3b.

Preparation of the coin-cell battery. Symmetric lithium cells (CR2032 coin cells) were assembled with two lithium electrodes (MTI Corp., 450 μm thick) and 1 M lithium hexafluorophosphate ($LiPF_6$) in ethylene carbonate:dimethyl carbonate (EC:DMC) (v:v = 1:1) as the electrolyte. Celgard 3501 was used as the separator. We subjected the cells to galvanostatic charging for 24 h at a current density of 1 mA cm^{-2}. A charging profile from one coin cell used is shown in Extended Data Fig. 7.

To plunge-freeze the samples, slush nitrogen was chosen to avoid detrimental interactions of the electrolyte with typical organic cryogens[37]. The slush was produced in the cryo-FIB workstation by vacuum-pumping liquid N_2 until it solidified. This enabled a higher cooling rate and reduced bubbling in the workstation. The coin cells were opened at the cryo-FIB workstation and the electrodes separated and immediately plunged into the slush nitrogen to preserve the electrolyte on the electrode. The frozen samples were then transferred into the preparation chamber attached to the cryo-FIB, typically sputter-coated with a 5–10-nm layer of metal (Pt or Au/Pd) to reduce charging, and transferred into the cryo-FIB chamber. Lift-out samples produced for cryo-STEM were transferred back to the workstation, where they were loaded into cryogenic sample storage boxes under liquid N_2 and transferred to a large liquid N_2 storage dewar. Cryo-TEM diffraction on lamellae produced by cryo-FIB lift-out showed that the electrolyte was frozen amorphously, and remained so throughout all of the preparation, transfer and characterization steps, as shown in Extended Data Fig. 8.

Three-dimensional reconstruction of cryo-FIB cross-sections. To reconstruct the three-dimensional dendrite structures, we used Avizo software (Thermo Fisher Scientific). Since the geometry of the FIB results in images of the cross-sections taken at oblique angles to the cross-section surface (the electron and ion columns are separated by 52° and the sample surface normal is positioned parallel to the ion beam for milling), the 30–50 individual images were aligned vertically by the position of the electrode surface and the appropriate length transformations were applied to the images to correct for the oblique viewing angle ($y = y'/\cos(\theta)$ and $z = z'/\sin(\theta)$, where y and z are the true depth and height of the object, respectively, y' and z' are the observed depth and height, and θ is the angle between the electrode surface normal and the electron beam). The dendrite and electrode structures were segmented by hand within each of the cross-sectional images, and these segmentations were connected in the perpendicular direction to reconstruct the three-dimensional structure.

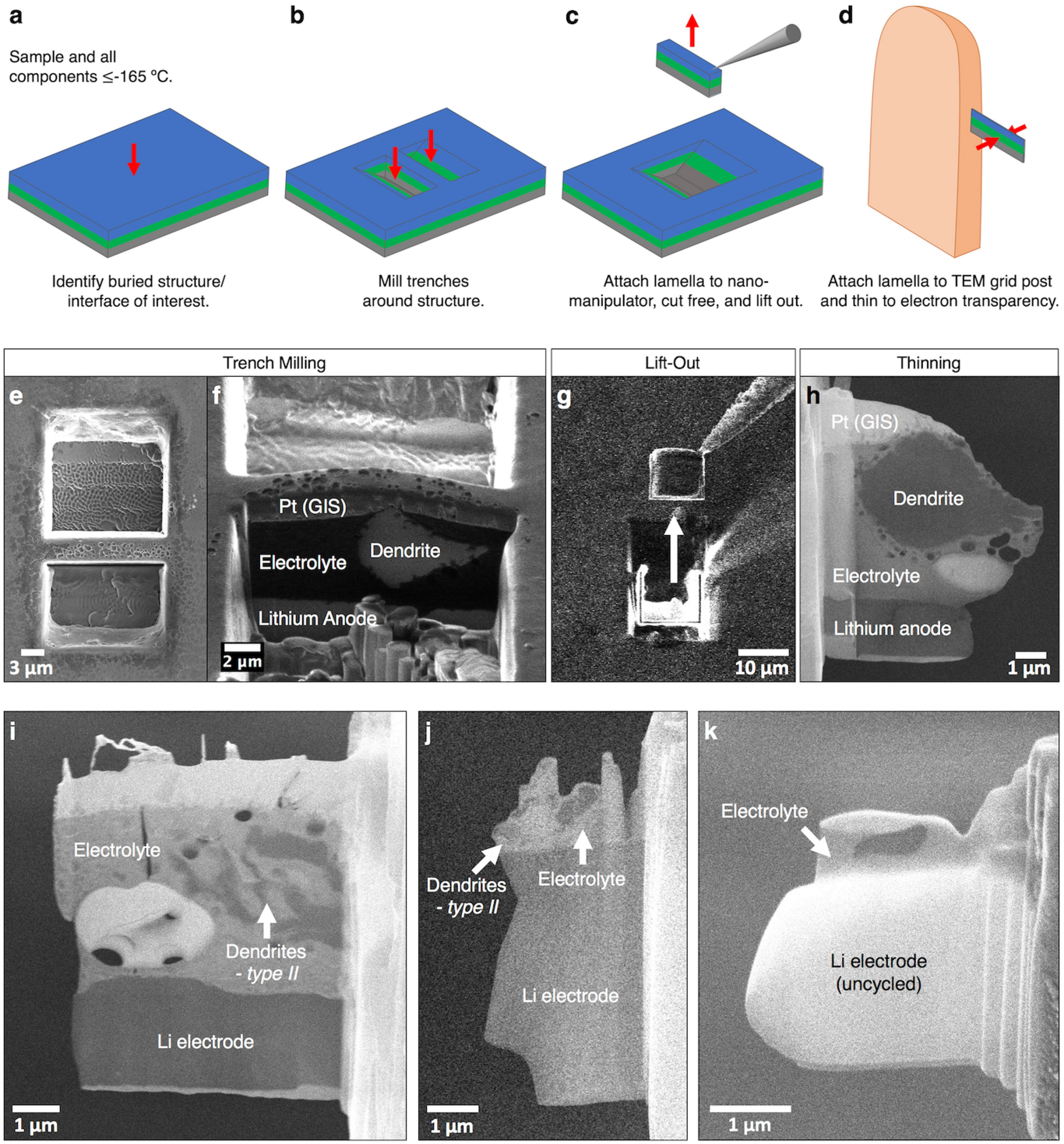
EELS map processing. The large field of view of some EELS maps results in an energy shift of the entire spectrum at different points in the map. To accurately

map edges and analyse edge fine structure across the field of view, the energy axis at each pixel was shifted to the proper location. Each map was acquired in DualEELS mode, with both low-loss and high-loss regions of the spectrum recorded. The low-loss maps included the zero-loss peak, the position of which was used to align the energy axis of the low- and high-loss regions of the spectra simultaneously, resulting in a flat energy surface across the map.

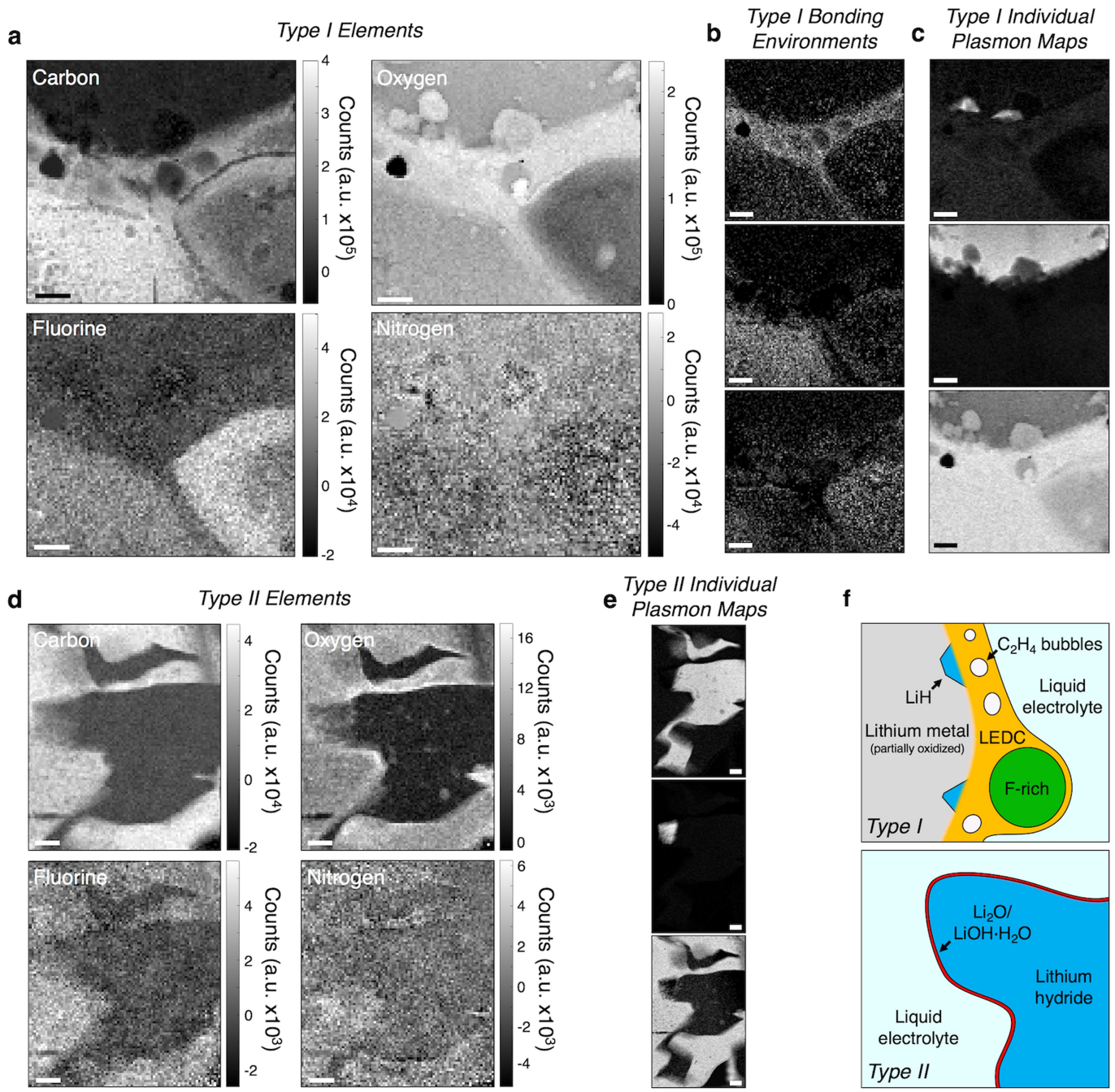
To map elemental distributions, standard background subtractions were performed using a linear combination of power laws fit and local background averaging with a full-width at half-maximum of five pixels to increase the background signal-to-noise ratio[38]. Energy windows wider than the fine structure at the edge onset were integrated for elemental mapping, to minimize the effects of spatially varying fine structure on the apparent elemental concentrations. The fine structure of the edges was analysed by multivariate curve resolution (MCR), which solves for a specified number of linearly independent spectral components in the data by means of a local minimization. A non-negativity constraint was imposed on the corresponding concentration profiles, since negative concentrations are not physical, but the spectra were not constrained. To improve the signal-to-noise ratio for the MCR process, the data was typically binned by four (spatially) before analysis. To display the spatial distribution of the resulting spectral components, we fitted them back to the original data using Matlab's QR solver, which takes advantage of QR factorization to minimize the residual of the equation $SC = D$. In our case, S is the matrix of spectral components returned by MCR, C is the matrix of concentrations to be solved for and D is the matrix of original data. Using the original unbinned data (for the type I maps) or the original data binned by two (for the type II maps), we produced maps from the concentration matrix C for the corresponding MCR spectral components, such as in Fig. 3.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

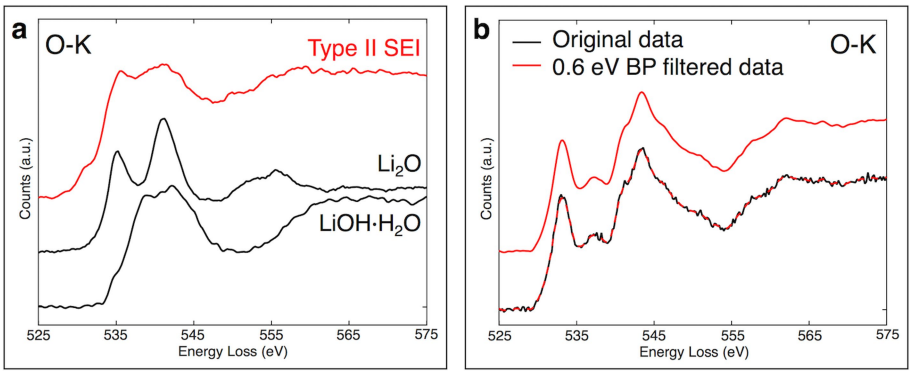
31. Tate, M. W. et al. High dynamic range pixel array detector for scanning transmission electron microscopy. *Microsc. Microanal.* **22**, 237–249 (2016).
32. Egerton, R. F., Li, P. & Malac, M. Radiation damage in the TEM and SEM. *Micron* **35**, 399–409 (2004).
33. Lin, F., Markus, I. M., Doeff, M. M. & Xin, H. L. Chemical and structural stability of lithium-ion battery electrode materials under electron beam. *Sci. Rep.* **4**, 5694 (2014).
34. Miki, T., Ikeya, M., Kondo, Y. & Kanzaki, H. Reflectance spectrum of lithium hydride at the Li K-absorption edge. *Solid State Commun.* **39**, 647–649 (1981).
35. Liu, D.-R. Electron energy loss spectroscopy of LiH with a scanning transmission electron microscope. *Solid State Commun.* **63**, 489–493 (1987).
36. Parades Mellone, O. A., Ceppli, S. A., Arneodo Larochette, P. P. & Stutz, G. E. Excitación de electrones K del Li a baja transferencia de momento por dispersión inelástica de rayos X en LiH. *Anal. Asoc. Fís. Argentina* **26**, 93–97 (2015).
37. Kesselman, E. et al. Cryogenic transmission electron microscopy imaging of vesicles formed by a polystyrene–polyisoprene diblock copolymer. *Macromolecules* **38**, 6779–6781 (2005).
38. Cueva, P., Hovden, R., Mundy, J. A., Xin, H. L. & Muller, D. A. Data processing for atomic resolution electron energy loss spectroscopy. *Microsc. Microanal.* **18**, 667–675 (2012).



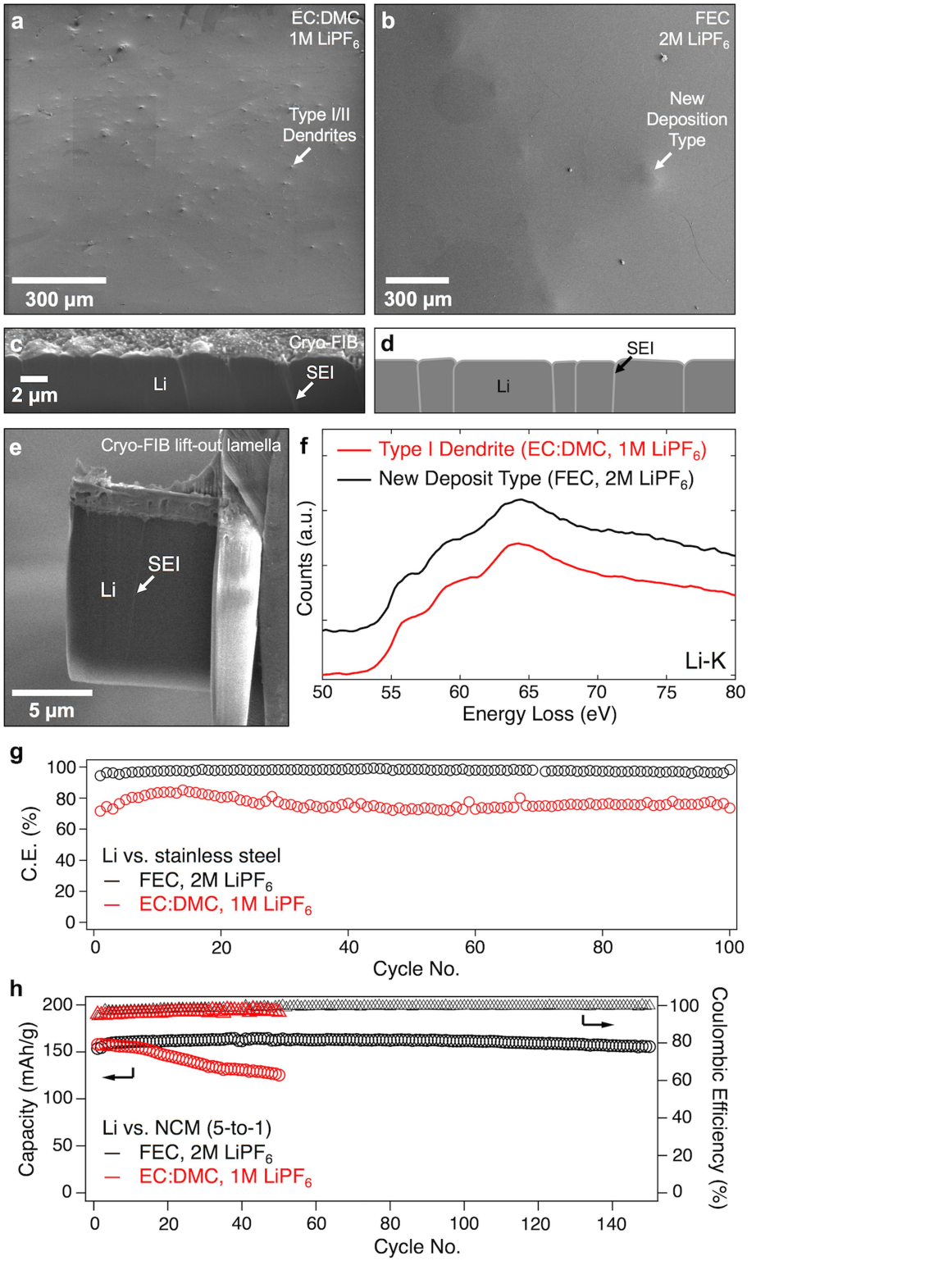
Extended Data Fig. 1 | Schematic and SEM images of the cryo-FIB lift-out sample preparation process, and examples of additional final lamellae. a, A buried structure or interface is identified for preparation, here a dendrite embedded in frozen electrolyte above the anode (indicated by the red arrow). In our coin-cell batteries, raised regions of electrolyte were used to localize buried dendrites. **b**, **e**, **f**, Trenches are then site-specifically milled around the site of interest, forming a vertical cross-sectional lamella containing the structure or interface. The sample is aligned in the microscope so that the electrode surface normal is parallel to the electron beam direction in **e**, and tilted by 52° to image the lithium anode–electrolyte interface and the electrolyte-embedded dendrite in **f**. **c**, **g**, A cooled nanomanipulator needle is then attached to the cryo-immobilized lamella by water vapour from a gas-injection system deposited as amorphous ice. The lamella is then cut free from the sample and lifted out. **d**, **h**, Finally, the lamella is attached to a TEM grid post with additional ice deposition, cut free from the nanomanipulator and thinned to electron transparency with the ion beam. **i**, **j**, Lamellae containing type II dendrites above lithium electrodes. The lamella in **i** contains a fluorine-rich structure as well. Different electrolyte thicknesses and milling parameters were used to prepare these lamellae, resulting in different final dimensions. **k**, A lamella produced from an uncycled electrode, used to obtain reference spectra. The increased signal of the uncycled electrode is due to different image-acquisition parameters, not a material difference.



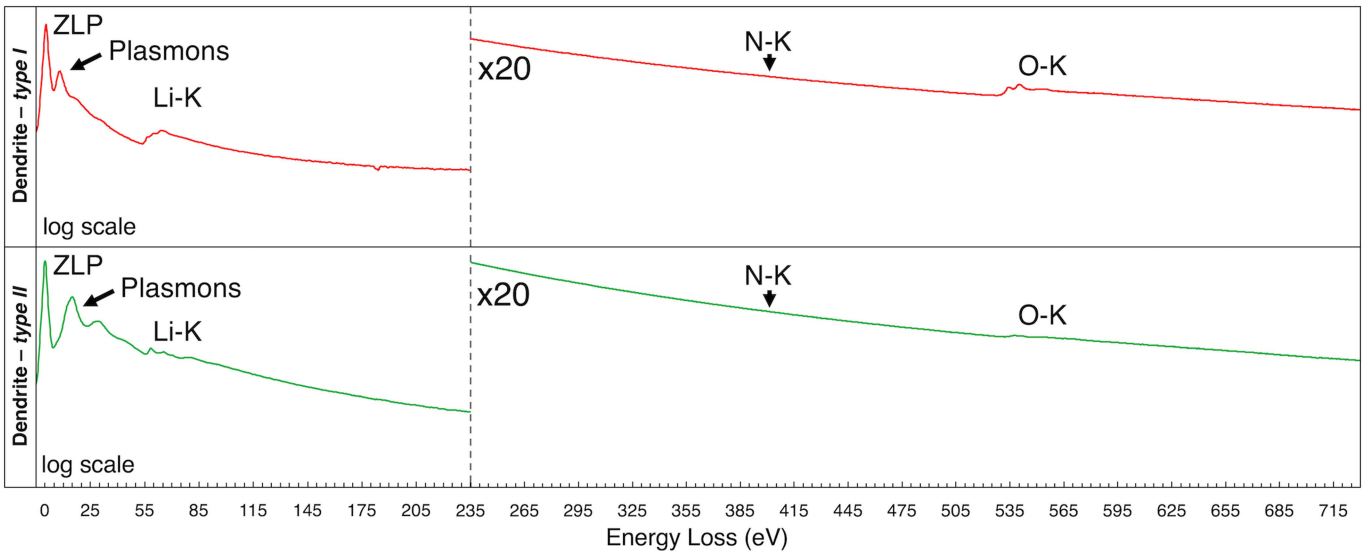
Extended Data Fig. 2 | Elemental maps of the regions near both types of dendrite surface, carbon-bonding environment maps resulting from fitting of MCR spectra back to original data, and corresponding summary schematics of both dendrite types and their SEI layers. Carbon, oxygen and fluorine are shown in a composite map in Fig. 2. **a**, **d**, Individual elemental maps showing the full count range, excluding 0.2% of high- and low-intensity outliers, make it clear that there is a substantial concentration of oxygen in the type I dendrite and very little in the type II dendrite, and that there is increased oxygen in the type I SEI compared to the electrolyte. In addition, essentially no fluorine is present in the type I SEI, and the large fluorine-rich structure contains a higher fluorine concentration than the electrolyte. Nitrogen maps are included as well, and largely show noise with little spatial dependence. Corresponding count scale bars are shown next to each map. **b**, Individual maps determined by MCR corresponding to the spectra primarily located at (and labelled in Fig. 3 as) the SEI, electrolyte and fluorine-rich structure (top to bottom), displaying the original counts. **c**, **e**, Individual plasmon maps determined by MCR for LiH, lithium and the electrolyte (top to bottom), displaying the original counts. **f**, Top, type I dendrites consist of partially oxidized lithium metal with small LiH regions at the surface, and have an extended SEI layer consistent with lithium ethylene dicarbonate (LEDC) that contains bubbles, probably from ethylene, a by-product of the SEI formation. Large fluorine-rich structures are often found near the dendrites. Bottom, type II dendrites consist of uniform LiH and have a compact $Li_2O/LiOH \cdot H_2O$ SEI layer. Although not depicted, fluorine-rich structures were also observed near type II dendrites. Scale bars, 300 nm.



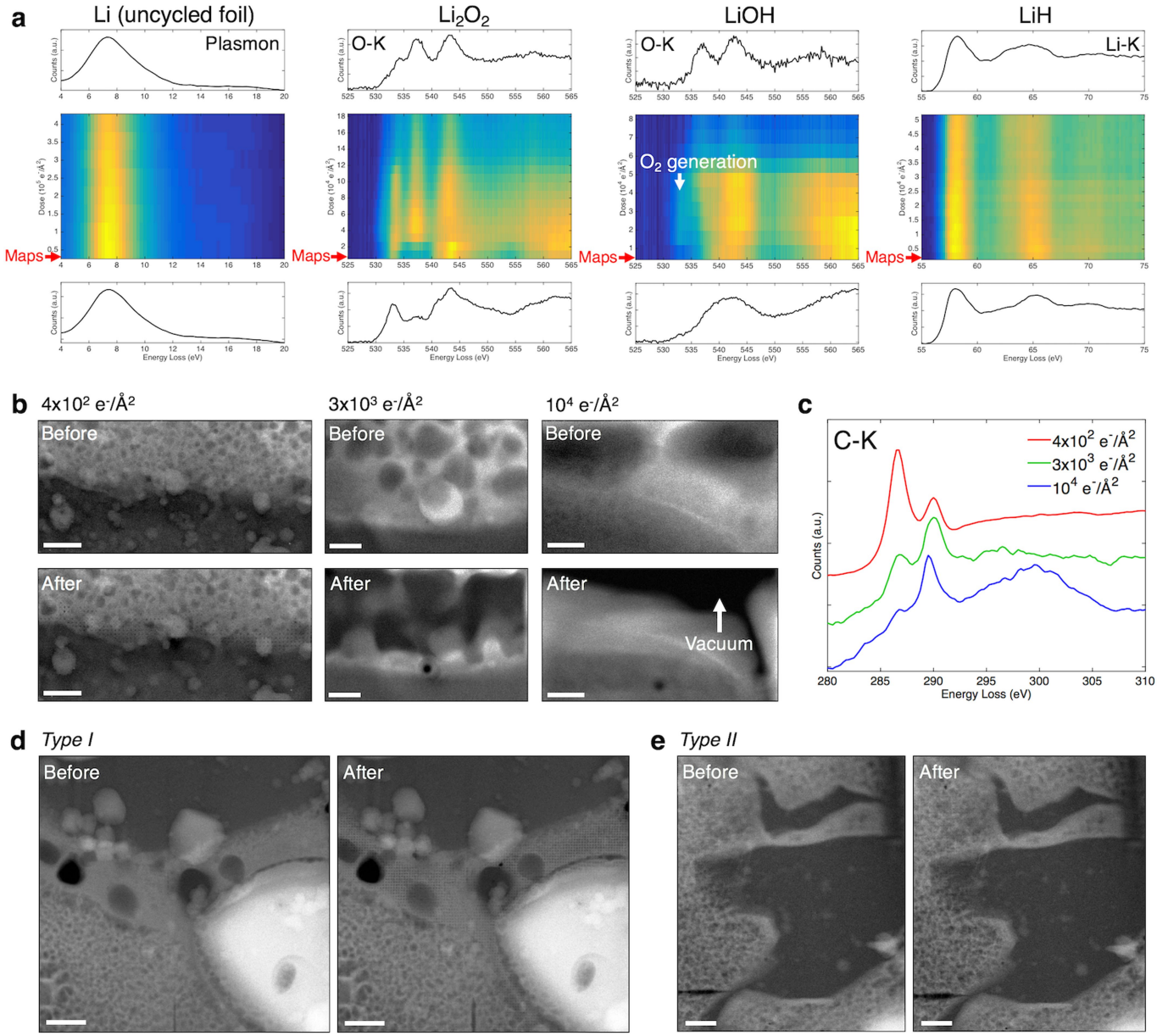
Extended Data Fig. 3 | Comparison of the type II SEI oxygen K-edge with reference spectra and an example of a bandpass-filtered spectrum. **a**, The O K-edge of the type II dendrite appears to be consistent with a combination of Li_2O and $LiOH \cdot H_2O$. Spectra are offset vertically for clarity. **b**, A 0.6-eV bandpass (BP) filter was applied to the O K-edge spectra acquired on the F20 to remove high-frequency noise. This preserved the main features of the edge while eliminating those below the energy resolution of the instrument.



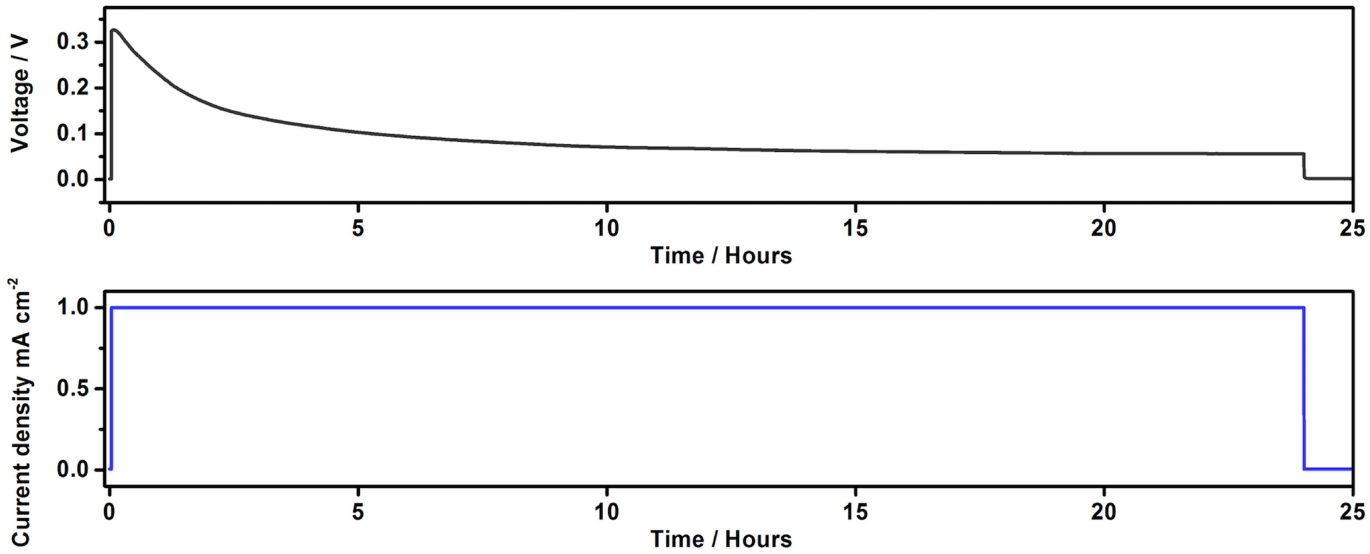
Extended Data Fig. 4 | Cryo-FIB, cryo-STEM EELS and electrochemical results comparing lithium deposition in cells using traditional and full-fluoride electrolytes. a, **b**, Cryo-FIB reveals that the dendrite density is much lower for the full-fluoride fluoroethylene carbonate (FEC) electrolyte (**b**) than with the traditional EC:DMC electrolyte (**a**). In the former case, nearly no LiH dendrites are present and the lithium deposition is modified, forming broad localized depositions. **c**, **d**, Cross-sections of these deposits reveal that they are composed of many smaller 'blocks' in contact, separated by SEI layers. **e**, **f**, A lamella of this type of deposition was prepared by cryo-FIB lift-out (**e**) and cryo-STEM EELS of the Li K-edge of the material revealed that it is composed of partially oxidized lithium metal (**f**), as was the type I dendrite in the traditional electrolyte. **g**, The Coulombic efficiency measured in a lithium versus stainless steel set-up using a constant current density of 1 mA cm^{-2} and capacity of 1 mAh cm^{-2} was greatly improved for the full-fluoride electrolyte compared to the traditional electrolyte. **h**, Cycling of a full cell comprising a lean lithium anode (50 µm) and a nickel manganese cobalt oxide (NCM) cathode (2 mAh cm^{-2}) with the full-fluoride electrolyte resulted in a substantial decrease in capacity fade and improved Coulombic efficiency over the traditional electrolyte. The discharge capacity is plotted on the left axis, whereas the Coulombic efficiency is on the right axis. The operating voltage range was 4.3 V to 3 V. In all figures, the red lines and symbols represent results for the EC:DMC, 1 M $LiPF_6$ electrolyte, whereas the black lines and symbols are for FEC, 2 M $LiPF_6$.



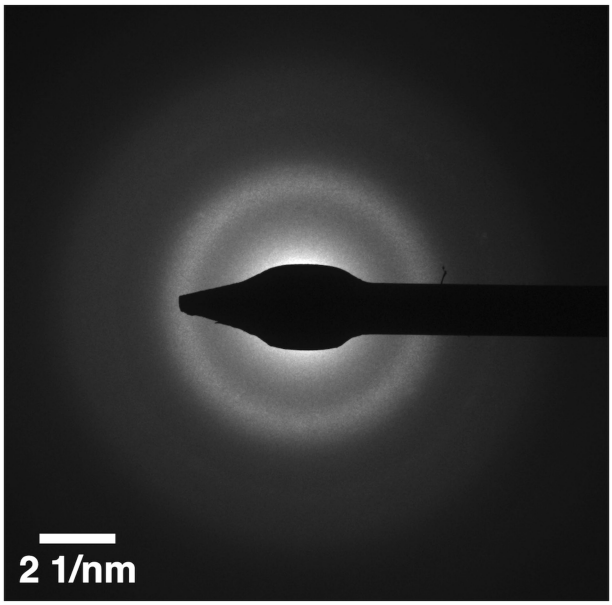
Extended Data Fig. 5 | Full spectra recorded from the dendrites (intensities on a logarithmic scale). The spectra show clear differences in the plasmons and Li K-edges, as well as a large difference in oxygen content between the type I and type II dendrites. The small amount of oxygen on the type II dendrite is probably due to water molecules adsorbed on the surface of the sample in the microscope vacuum, which would typically react with materials such as lithium or sodium at room temperature. No nitrogen was present in either dendrite, confirming that no reaction with nitrogen in the air or liquid N_2 had occurred.



Extended Data Fig. 6 | Example damage series profiles and initial/final spectra taken for lithium materials relevant to this study over a range of doses at which damage occurs, dark-field cryo-STEM images of various types of damage induced in a frozen organic electrolyte at different doses with corresponding spectra, and before and after images of the regions in which the EELS maps in the main text were taken. All spectra were recorded at cryogenic temperatures. **a**, We found all oxide materials convert to Li_2O under large doses. Li_2O and LiH are primarily affected by mass loss, with no substantial changes in fine structure. The maps presented in the main text were acquired at doses lower than the dose indicated by the red arrows shown at the bottom of the plots, of the order of $10^2\ e^-\ Å^{-2}$. **b**, **c**, While some structural modification of the electrolyte material was present at low doses, probably due to liberation of hydrogen, a dose greater than $10^3\ e^-\ Å^{-2}$ was required for substantial mass loss and modification of spectral fine structure. At $3 \times 10^3\ e^-\ Å^{-2}$, approximately 50% of the material remained after the map, as determined by the ADF signal. At $10^4\ e^-\ Å^{-2}$, the material was completely removed in some areas, but the carbonate portion of the molecule remained. Doses applied during acquisition of the maps in the main text were less than the lowest dose shown here. Spectra are offset vertically for clarity. **d**, **e**, In the maps displayed in the main text, small structural changes were observed in the organic materials, which is expected given our damage analysis. This is probably due to liberation of hydrogen from the molecules, which occurs at low dose. The fine structure is not greatly affected until approximately an order of magnitude higher dose than was applied during these maps, which was of the order of $10^2\ e^-\ Å^{-2}$. Scale bars, 200 nm, 30 nm and 60 nm (**b**, left to right), 300 nm (**d**, **e**).



Extended Data Fig. 7 | Charging profile from a symmetric lithium coin cell. A constant current of 1 mA cm^{-2} was applied to the cells for 24 h (bottom). The resulting voltage profile from one of the coin cells used is shown in the top panel.



Extended Data Fig. 8 | Amorphous diffraction pattern of the electrolyte recorded in a cryo-lamella produced by cryo-FIB lift-out. Cryo-TEM diffraction of the electrolyte on samples produced by cryo-FIB lift-out shows that it is frozen amorphously and does not recrystallize at any point during the preparation, storage, transfer or characterization.

Extended Data Table 1 | Comparison of the properties of type I and II dendrites

Property	Dendrite Type I	Dendrite Type II
Diameter	~1-10 μm	~100s of nm
Morphology	Low curvature	Tortuous
Electrode Contact Area	~10 μm^3	~0.1 μm^3
Composition	Lithium metal, slightly oxidized (Li_2O).	LiH
Distribution of Materials	Slightly oxidized lithium with small LiH regions on surface.	Uniform LiH. Pure/slightly oxidized Li particle on tip.
SEI Thickness	~300-500 nm	~20 nm
SEI Elements Detected	Li, C, O	Li, O
SEI Composition	Extended SEI consistent with lithium ethylene dicarbonate. Large bubbles within likely originate from ethylene gas.	Li_2O and $LiOH{\cdot}H_2O$

Extended Data Table 2 | Threshold electron doses and primary damage mechanisms observed for relevant materials

Material	Damage Dose ($e^-/Å^2$)	Primary Damage Mechanism
Li (uncycled foil)	10^5	Mass loss/Minimal Li_2O formation
Li_2O	$>10^5$	Mass loss
Li_2O_2	$>10^4$	Mass loss/Li_2O conversion
LiOH	10^4	O_2 generation/Mass loss/Li_2O conversion
LiH	$>10^4$	Mass loss
EC:DMC - 1M $LiPF_6$	$<10^2$ $>10^3$ 10^4	Slight structural modifications Mass loss/Initial fine structure changes Large mass loss/Conversion to Li_2CO_3

All of the damage thresholds listed are for cryogenic samples under a 300-kV electron beam. The primary damage mechanism of lithium is mass loss. While a very small amount of oxide can form, this is much less than the initial oxide impurities in an uncycled electrode, and occurs at more than 10^5 e^- $Å^{-2}$. Li_2O_2 and LiOH both initially suffer mass loss at or above 10^4 e^- $Å^{-2}$ and subsequently begin converting to Li_2O at higher doses, whereas Li_2O suffers only mass loss and no change in fine structure. LiH is stable to above 10^4 e^- $Å^{-2}$, with mass loss occurring above this and little to no oxidation. The electrolyte material damages differently at different doses, with slight structural modifications occurring at low doses and mass loss and changes in fine structure occurring above 10^3 e^- $Å^{-2}$. At 10^4 e^- $Å^{-2}$, a substantial portion of the mass is lost, leaving behind the carbonate portion of the solvent molecules, the fine structure of which is still intact at this dose.

LETTER

https://doi.org/10.1038/s41586-018-0391-9

Building C(sp^3)-rich complexity by combining cycloaddition and C–C cross-coupling reactions

Tie-Gen Chen[1,3], Lisa M. Barton[1,3], Yutong Lin[1,3], Jet Tsien[1], David Kossler[1], Iñaki Bastida[1], Shota Asai[1], Cheng Bi[1], Jason S. Chen[1], Mingde Shan[2], Hui Fang[2], Francis G. Fang[2], Hyeong-wook Choi[2], Lynn Hawkins[2], Tian Qin[1] & Phil S. Baran[1]*

Prized for their ability to rapidly generate chemical complexity by building new ring systems and stereocentres[1], cycloaddition reactions have featured in numerous total syntheses[2] and are a key component in the education of chemistry students[3]. Similarly, carbon–carbon (C–C) cross-coupling methods are integral to synthesis because of their programmability, modularity and reliability[4]. Within the area of drug discovery, an overreliance on cross-coupling has led to a disproportionate representation of flat architectures that are rich in carbon atoms with orbitals hybridized in an sp^2 manner[5]. Despite the ability of cycloadditions to introduce multiple carbon sp^3 centres in a single step, they are less used[6]. This is probably because of their lack of modularity, stemming from the idiosyncratic steric and electronic rules for each specific type of cycloaddition. Here we demonstrate a strategy for combining the optimal features of these two chemical transformations into one simple sequence, to enable the modular, enantioselective, scalable and programmable preparation of useful building blocks, natural products and lead scaffolds for drug discovery.

Retrosynthetic chemical analysis is built upon the strategic identification of the reactions (transforms) that offer the greatest potential to simplify target preparation (the 'T-goal')[3]. To this end, the capacity of cycloadditions to generate complex ring systems and multiple bonds with precise stereochemical control is unmatched. By contrast, C–C cross-couplings such as Heck, Suzuki and Negishi reactions are capable of making only one bond at a time (most often between flat sp^2 systems)[4], yet are the most used C–C bond forming methods in the patent literature[7]. To understand this phenomenon, it is worthwhile to compare the features of these two diverse reaction classes (Fig. 1a). Cycloadditions form two new sigma bonds, generally through concerted pathways that follow precise rules for predicting stereo- and regiochemistry. Starting with specialized building blocks that are designed to enable the reaction to take place cleanly, this process rapidly accesses complexity[8]. On the other hand, C–C cross-couplings form one new sigma bond between easily accessible building blocks, using a transition-metal catalyst to reliably and controllably produce new connections. What C–C cross-coupling lacks in terms of complexity generation, it makes up for in terms of sheer reliability and modularity.

These distinct features are illustrated by the syntheses of the alkaloid epibatidine (**1**)[9] and the commercial antihypertensive medicine Cozaar (**2**)[10] (Fig. 1a). Epibatidine, coveted for its pronounced analgesic properties, has been prepared by total synthesis more than 60 times (see Supplementary Information for a complete listing). Of these syntheses, at least 31 have used cycloaddition chemistry as their key ring-constructing step. The vast majority of these approaches involve formation of the bridged pyrrolidine core, followed by stepwise, and often lengthy, pyridine incorporation. Meanwhile, medicinal agent **2**, bereft of any stereogenic centres or topological complexity, was heralded as one of the first examples of a 'blockbuster' drug. Both its discovery and its eventual manufacture used C–C cross-couplings (Ullmann and Suzuki reactions) for the key bond cleavage[11,12]. This facilitated both a convergent assembly and a modularity that permitted the rapid exploration of hundreds of analogues.

Here, we sought to combine the innate complexity generation of cycloaddition with the simplicity and modularity of C–C coupling. When applied to structures such as **1**, this strategy will permit the rapid generation of analogues, and when applied to medicinal programs such as **2**, it will allow for the rapid exploration of otherwise challenging complex chemical space.

Building block **3**, of hypothetical value in medicinal chemistry, represents the manifestation of this idea (Fig. 1b). Although its structure would seem to suggest that it could be formed through a Diels–Alder reaction, the relevant synthetic building blocks, structures **4** and **5**, are not electronically matched and therefore one would expect no reaction to take place. Even if a workable enantioselective Diels–Alder reaction could be invented to achieve this transform, the strategy would suffer from a lack of modularity. In order to solve this problem, one could envisage using a hypothetical vicinal dihaloethylene (**6**) in place of **5**, with a fumarate-type dienophile such as **7** serving as a viable synthetic equivalent. The favourable matched electronics of the dienophile should allow for a facile Diels–Alder reaction and subsequent radical cross-coupling (RCC).

To address the enantioselectivity challenge, a combination of transforms was proposed (Fig. 1c). Maleic anhydride was chosen as a surrogate for the hypothetical chiral (pseudo)dihalide **6** given its inexpensive nature, ready participation in most cycloaddition modes ([2+1], [2+2], [3+2], [4+2]), and known desymmetrization through chiral Lewis-base-mediated alcoholysis. Our sequence for generating complexity in a modular fashion involves five simple steps: first, cycloaddition to build a scaffold; second, desymmetrization to set absolute stereochemistry; third, RCC to install a new C–C bond; fourth, hydrolysis; and fifth, RCC to forge another new C–C bond. The known versatility of RCC enables a range of functionalities to be installed, from aryl[13] and heteroaryl[14] systems to alkenes[15], alkynes[16] and alkyl groups[17]. We describe here the preparation of more than 80 synthetic examples and multiple applications, covering a range of natural products (including the synthesis of **1**) and real-world examples from industrial settings.

Studies commenced with Diels–Alder [4+2] cycloaddition (Fig. 2a), by which a large variety of enantiomerically enriched scaffolds could be produced in a simple modular fashion. First, scaffolds $\mathbf{A_1}$–$\mathbf{A_6}$ ($\mathbf{A_2}$ being a Diels–Alder adduct and all others being derived from Diels–Alder/hydrogenation) were desymmetrized using Deng's conditions, with either quinine or quinidine used as the Lewis base to deliver mixed acids/esters in excellent enantioselectivity[18]. Next, the mono-acid substrate was subjected to successive Negishi[13,15–17] and Suzuki[14] type RCC reactions, depending on starting-material availability or individual preferences. In this manner, some of the most simple and inexpensive Diels–Alder adducts known ($\mathbf{A_1}$ costs US$38.54 per mol and $\mathbf{A_2}$ is US$9.52 per mol) can be transformed into enantioenriched products (**11**–**63**) of high value. Indeed, none of these products can at present be prepared using a Diels–Alder reaction (racemic or enantioenriched),

[1]Department of Chemistry, The Scripps Research Institute (TSRI), La Jolla, CA, USA. [2]Integrated Chemistry Engine, Eisai AiM Institute, Andover, MA, USA. [3]These authors contributed equally: Tie-Gen Chen, Lisa M. Barton, Yutong Lin. *e-mail: pbaran@scripps.edu

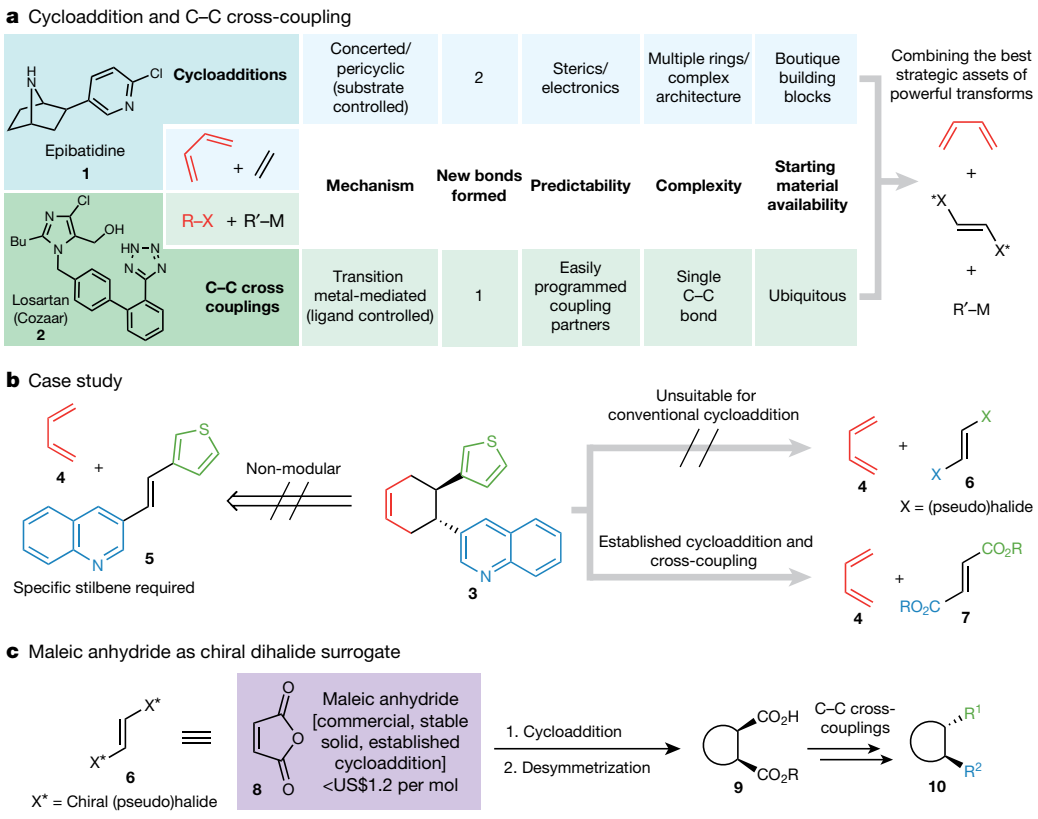


Fig. 1 | Combining the logic of cycloaddition and C–C cross-coupling. a, Comparison of cycloaddition and C–C cross-coupling. Bu, butyl; R–X, electrophile; R′–M, nucleophile. **b**, Case study: retrosynthetic analysis of enantiopure building block **3**. This building block could in theory be formed through a Diels–Alder reaction, but the structures that would be used for this, **4** and **5**, are not electronically matched and no reaction would occur; moreover, the strategy would not result in any modularity. Instead a hypothetical vicinal dihaloethylene (**6**) could be used; however, this is unsuitable for conventional cycloaddition. **7** is a viable synthetic equivalent. **c**, Maleic anhydride can be used as an available and modular chiral dihalide surrogate.

because in most cases the reaction would require both an electron-rich diene and an electron-rich dienophile and therefore is electronically unfavoured. Furthermore, controlling the chemoselectivity when there are multiple alkenes present in the dienophile is challenging using traditional Diels–Alder chemistry; however, with our method, diverse alkenes and alkynes can be easily installed post-cyclization, with excellent isomeric and geometrical purity (accessing **12**, **15**, **19**, **21**, **25** and **31**). Typically, scaffolds derived from Diels–Alder adducts $\mathbf{A_1}$–$\mathbf{A_6}$ have a clear retrosynthetic signature, wherein the electron-withdrawing groups on the dienophile have been homologated, alkylated, or degraded. Therefore, our approach allows access to the previously unexplored chemical space of electron-rich, chiral, 1,2-disubstituted Diels–Alder scaffolds. Using the tactical combination outlined above, a virtually limitless array of substituents is now easily accessed, including halogenated arenes, Lewis-basic heterocycles, α,β-unsaturated carbonyl groups, cyclopropanes, cyclobutanes, sulfur moieties, and alkenes. In the case of *exo* and *endo* isomers $\mathbf{A_4}$ and $\mathbf{A_5}$ (both commercially available), the sequences converged to an identical product using the same coupling partners (**27** could be prepared from $\mathbf{A_4}$ or $\mathbf{A_5}$).

As shown in Fig. 2b–d, the strategy outlined above could also be applied to [3+2], [2+2] and [2+1] cycloadditions. The vast scope observed with Diels–Alder chemistry was also seen in these cases, accessing substituents such as substituted olefins, terminal alkynes, homoallyl groups, and heterocycles. Pyrrolidine-containing systems could be derived from simple building blocks $\mathbf{B_1}$ and $\mathbf{B_2}$ (accessed through dipolar cycloaddition)[19] to furnish **64**–**82** with high enantiomeric excess. Quick access to pyrrolidine-containing drug scaffolds is useful: historically these have been extremely relevant to medicinal chemists, with around 40 pharmaceuticals containing this motif[20]. As with the Diels–Alder adducts, none of these structures has been accessed before. Therefore, our approach serves as a modular entry to chiral variants of these coveted scaffolds.

In a similar vein, enantioenriched cyclopentenes were easily produced from palladium-catalysed, trimethylenemethane-based, formal [3+2] cycloaddition adduct $\mathbf{B_3}$[21]. Such structures are highly challenging to access in any other way. Scaffolds $\mathbf{C_1}$ and $\mathbf{C_2}$, representing entry to [2+2]-derived systems, could be similarly processed. Modular access to enantioenriched cyclobutanes is important and useful given the strict electronic requirements for photochemical cycloaddition and the documented challenge in achieving general asymmetric induction[22]. Access to 1,2-disubstituted cyclobutanes (**83**–**86**, derived from $\mathbf{C_1}$) compares favourably with photochemical approaches to such systems that frequently give inseparable racemic mixtures of regiomers and diastereomers. Furthermore, access to tetrasubstituted, chiral cyclobutanes is highly useful, as multiple families of dimeric and pseudodimeric cyclobutane natural products contain such structural motifs[23]. Structures such as **87**, which would be otherwise extremely difficult to access through conventional photochemistry, can be easily prepared in enantioenriched manner from maleic anhydride heterodimerization adducts such as $\mathbf{C_2}$. Finally, the strategy outlined above when applied to [2+1] cycloadditions using scaffolds $\mathbf{D_1}$ and $\mathbf{D_2}$ is a major departure from the common retrosynthetic logic normally applied to such structures. Conventional approaches, usually involving late-stage cyclopropanation of an olefin, suffer from lack of enantiocontrol in the absence of directing groups or specialized carbenoid donors and complex catalysts[24]. Structures **94** and **95** ($\mathbf{D_2}$-derived) are particularly illustrative of this fact: it would be extremely challenging to access either of these in an enantioselective fashion with current synthetic technology (cyclopropanation or C–C cross-coupling). As testament to the power of this strategy to access diverse libraries, we synthesized an additional 48 enantioenriched compounds in a similar manner (see Extended Data Figs. 1 and 2 for details).

To further demonstrate the potential of this approach to simplify synthesis, we present six applications in the total synthesis of natural products and in both early- and late-stage medicinal chemistry programs (Fig. 3a–f). As mentioned above, alkaloid **1** has been a popular target of both academic and industrial scientists. Through the application of cycloaddition and cross-coupling, the native carboxylate required for the Diels–Alder reaction can be used directly to produce epibatidine (**1**) in five steps (for optimization and in-depth analysis of previous approaches, see Supplementary Information) with a 38% gram-scale overall yield

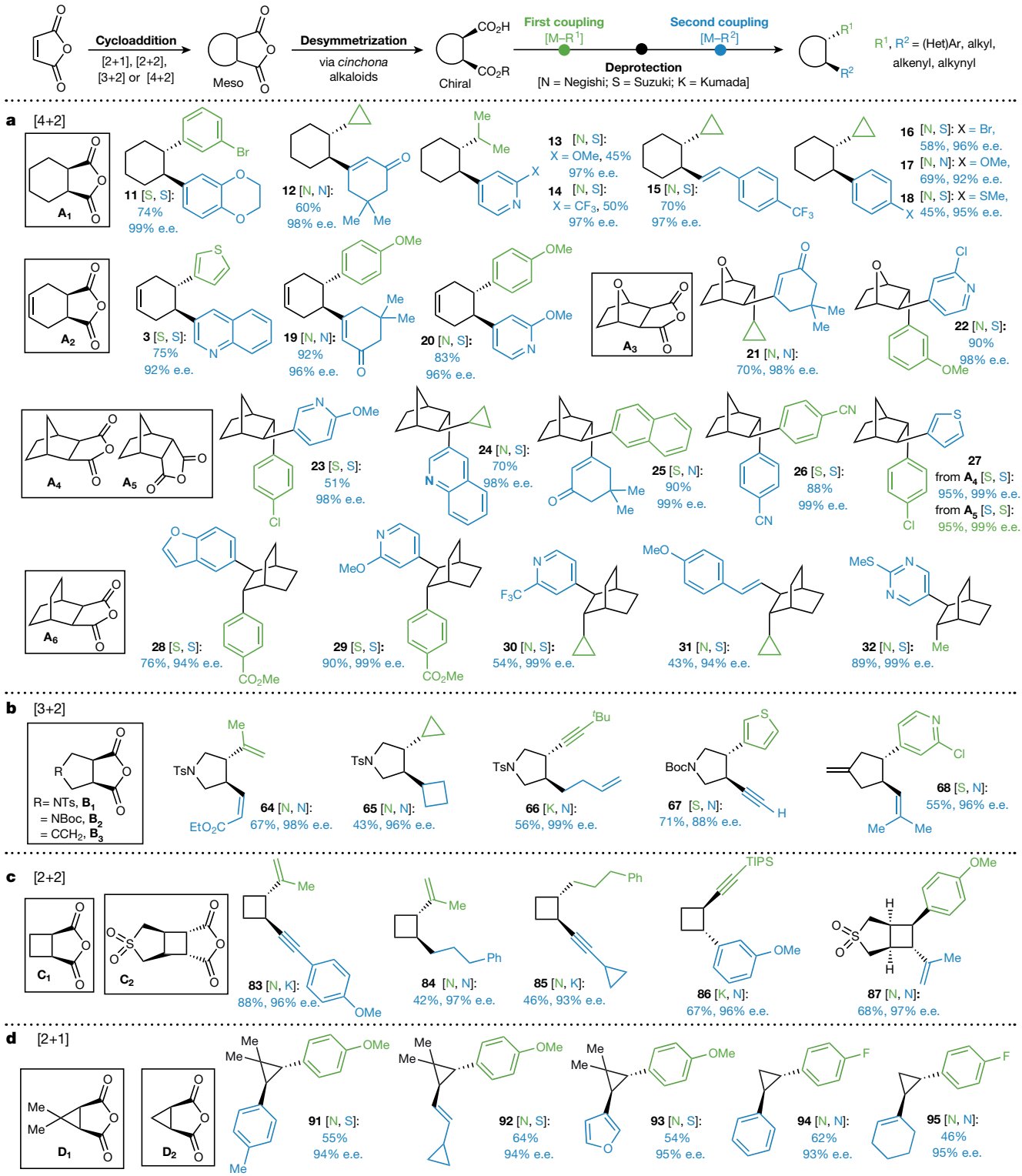


Fig. 2 | Substrate scope of combining cycloaddition and C–C cross-coupling. a–d, The cycloaddition component is shown in black, the first cross-coupling in green and the second cross-coupling in blue. The yield and enantiomeric excess (e.e.) refer to the second cross-coupling. Besides compound **89** (diastereromeric ratio (d.r.), see Extended Data Fig. 2), excellent diastereoselectivity (d.r. greater than 10/1) was observed in all cross-couplings. See Extended Data Figs. 1 and 2 for complete substrate scope and Supplementary Information for synthetic details. X-ray structure data are available for compounds **11**, **19**, **23**, **25**, **28**, **65** and $\mathbf{C_2}$. N, Negishi cross-coupling; S, Suzuki cross-coupling; K, Kumada cross-coupling; Boc, *tert*-butyloxycarbonyl; TIPS, triisopropylsilyl; Ts, tosyl.

(Fig. 3a). It is worth noting that the key decarboxylative cross-coupling takes place with 95% isolated yield (gram-scale, 72% yield).

Saphris (asenapine, **102**; Fig. 3b), an antipsychotic approved by the US Food and Drug Administration (FDA), is currently marketed as a racemic mixture (although the (+)-isomer has superior pharmacokinetic properties)[25]. This near-symmetric molecule has been challenging to prepare enantioselectively, as the two aromatic systems differ only in one chlorine substituent. It is therefore hard to envisage a cycloaddition that could be rendered enantioselective for the preparation of **102**, and only one enantioselective approach has been reported

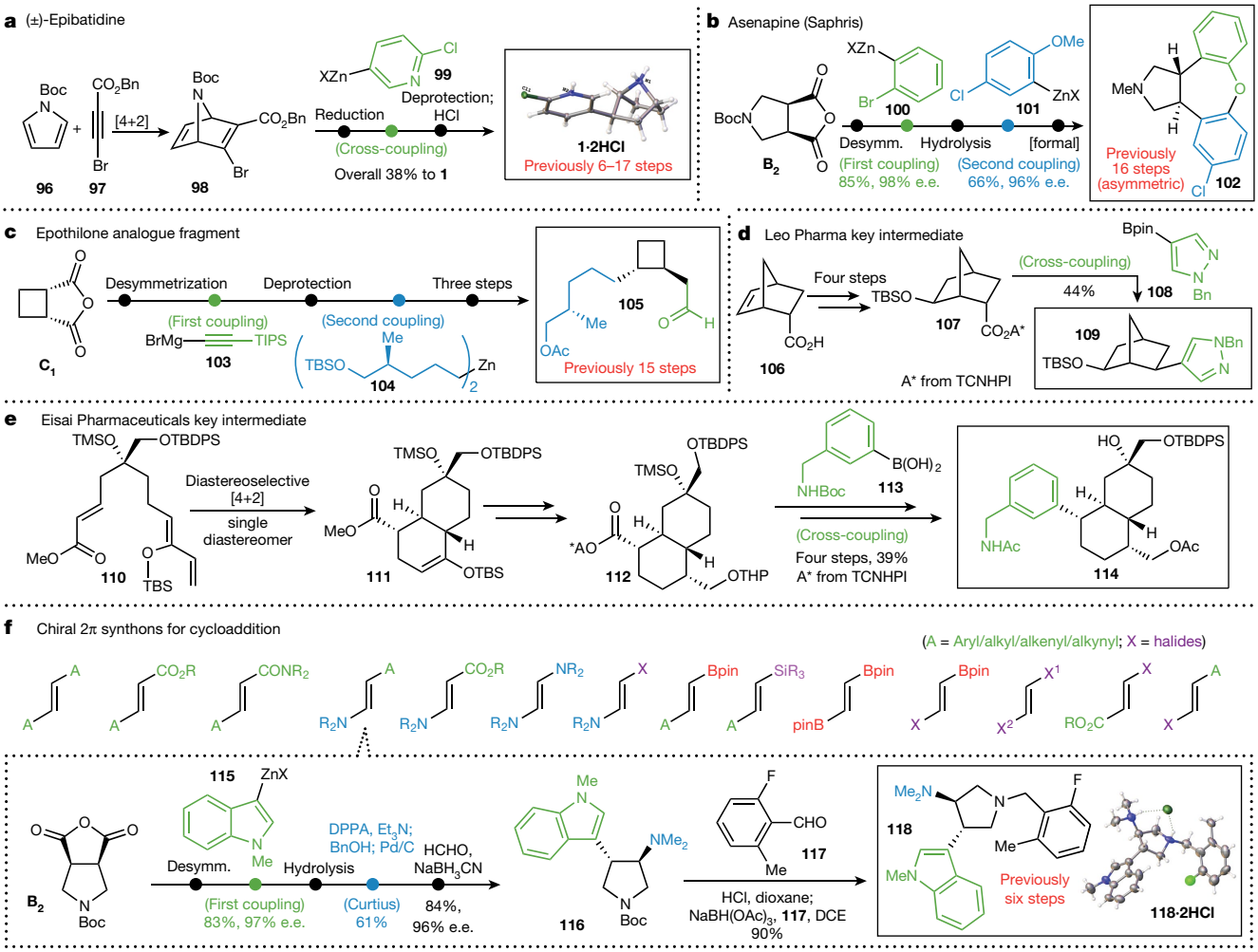


Fig. 3 | Applications of combining cycloaddition and C–C cross-coupling. a, Gram-scale synthesis of (±)-epibatidine. **b**, Asymmetric synthesis of asenapine (Saphris). **c**, Modular synthesis of intermediate **105**, a fragment used in the synthesis of a cyclobutyl-containing analogue of epothilone. **d**, Synthesis of a Leo Pharma key intermediate. **e**, Synthesis of an Eisai Pharmaceuticals key intermediate. **f**, New chiral 2π synthons for cycloaddition: application to the modular asymmetric synthesis of an inhibitor of the EED protein. Excellent diastereoselectivity (d.r. greater than 10/1) was observed in all cross-couplings. See Supplementary Information for full synthetic details and schemes. X-ray structure data are available for compounds **1·2HCl**, **107** and **118·2HCl**. Boc, *tert*-butyloxycarbonyl; Pin, pinacol group; TBDPS, *tert*-butyldiphenylsilyl; TBS, *tert*-butyldimethylsilyl; TCNHPI, tetrachloro-*N*-hydroxyphthalimide; THP, tetrahydropyranyl; TIPS, triisopropylsilyl; TMS, trimethylsilyl.

in 16 steps[26]. However, **102** can be prepared in formally six steps, with complete enantiocontrol, from $\mathbf{B_2}$, through a strategy that could also be used to make an array of near-symmetric analogues.

Epothilone—a famous natural product that inhibits the dynamics of cellular microtubules and inspired the FDA-approved medicine Ixabepilone[27]—has been the subject of numerous synthetic studies and analogue campaigns[28]. Intermediate **105** (Fig. 3c) has been used during Nicolaou's synthesis of a cyclobutyl-containing analogue of epothilone, but required 15 steps (24% overall yield) to be prepared in enantioenriched form from $\mathbf{C_1}$, via enzymatic desymmetrization and a series of carefully choreographed homologations[28]. Using the same starting material, $\mathbf{C_1}$, we have prepared intermediate **105** in only eight steps, through sequential desymmetrization, RCC with alkyne **103**, hydrolysis, RCC with alkyl zinc **104**, protecting-group exchange, and finally hydroboration/oxidation.

Differentially substituted norbornene rings have been shown to be useful phenyl bioisosteres in medicinal chemistry; however, their broad implementation is hindered by a lack of methods for their rapid modular construction (unlike the construction of aryl systems)[29]. In collaboration with Leo Pharma, we prepared a key target for an ongoing program (**109**; Fig. 3d) from Diels–Alder adduct **106** via alkene hydration and RCC with the pyrazole-boronic ester **108** in 44% yield. It is worth noting that this is, to our knowledge, the first report of a RCC reaction using a bis(pinacolato)diboron (Bpin) derivative rather than a boronic acid. This advance was achieved using an in situ prepared ate complex, prepared by adding one equivalent of *n*-BuLi relative to the aryl-Bpin donor[30]. We also used this modification of our Suzuki–RCC protocol to prepare cyclopropanes **91**, **92** and **93**.

Finally, an ongoing program at Eisai Pharmaceuticals necessitated the preparation of complex scaffold **114** (Fig. 3e), wherein the key structure–activity relationship to be explored resided at the aryl (green) portion of the molecule. This is a particularly powerful application of the strategy outlined herein, because a carboxylate—needed to achieve the diastereoselective Diels–Alder reaction to construct the decalin framework—served as a gateway for the exploration of chemical space at the desired position. Thus, an asymmetric Diels–Alder reaction using Corey's oxazaborolidine[31] catalyst, followed by functional group manipulations (see Supplementary Information), led to intermediate **112**, which could be cross-coupled with boronic acid **113** in a particularly challenging context to deliver **114** and enable biological follow-up.

The strategy outlined here could be useful for more than just forging C–C linkages through cross-coupling. The incorporation of classic nucleophilic functionalization (in ketone synthesis) and decarboxylative functionalization (in the Hunsdiecker reaction and in Curtius

and Wolff rearrangements) of intermediate adducts opens up innumerable possibilities for diversification, rendering access to previously inaccessible building blocks (Fig. 3f). As an example, we used a chiral 2π enamine synthon to prepare **118**, an inhibitor of the embryonic ectoderm development (EED) protein ($K_i = 4$ nM), which was previously prepared by AbbVie[32] through a non-modular, racemic [3+2] approach in 1.9% overall yield. Our approach began with intermediate *meso* scaffold $\mathbf{B_2}$ and used RCC followed by Curtius rearrangement, yielding optically pure **118** in 13% overall yield.

The advance that we have described is largely strategic in nature, and thus the underlying limitations are tied mainly to individual parameters of a particular cycloaddition and ensuing RCC reactions. That said, *cis*-1,2-disubstituted products are not currently accessible unless downstream isomerization reactions are pursued, which need to be evaluated on a case-by-case basis. Although ligand-controlled RCC reactions might eventually address this issue, in its present form this platform for modular molecular assembly holds great promise for accessing new areas of chemical space. The combination of classic cycloaddition chemistry with newly emerging radical C–C coupling offers a powerful way to repurpose the most classic skeleton-building reactions of organic synthesis, to simplify the enantioselective preparation of building blocks, natural products and medicines.

Data availability

Metrical parameters for the structures of **11**, **19**, **23**, **25**, **28**, **44**, **45**, **49**, **65**, **C2**, **107**, **1·2HCl** and **118·2HCl** are available free of charge from the Cambridge Crystallographic Data Centre (CCDC) under reference numbers 1837572, 1837575, 1837577, 1837578, 1837579, 1937573, 1837574, 1937576, 1821880, 1821878, 1821879, 1825177 and 1838237, respectively. Data are available from the corresponding author on reasonable request.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0391-9

Received: 23 February 2018; Accepted: 31 May 2018;
Published online 30 July 2018.

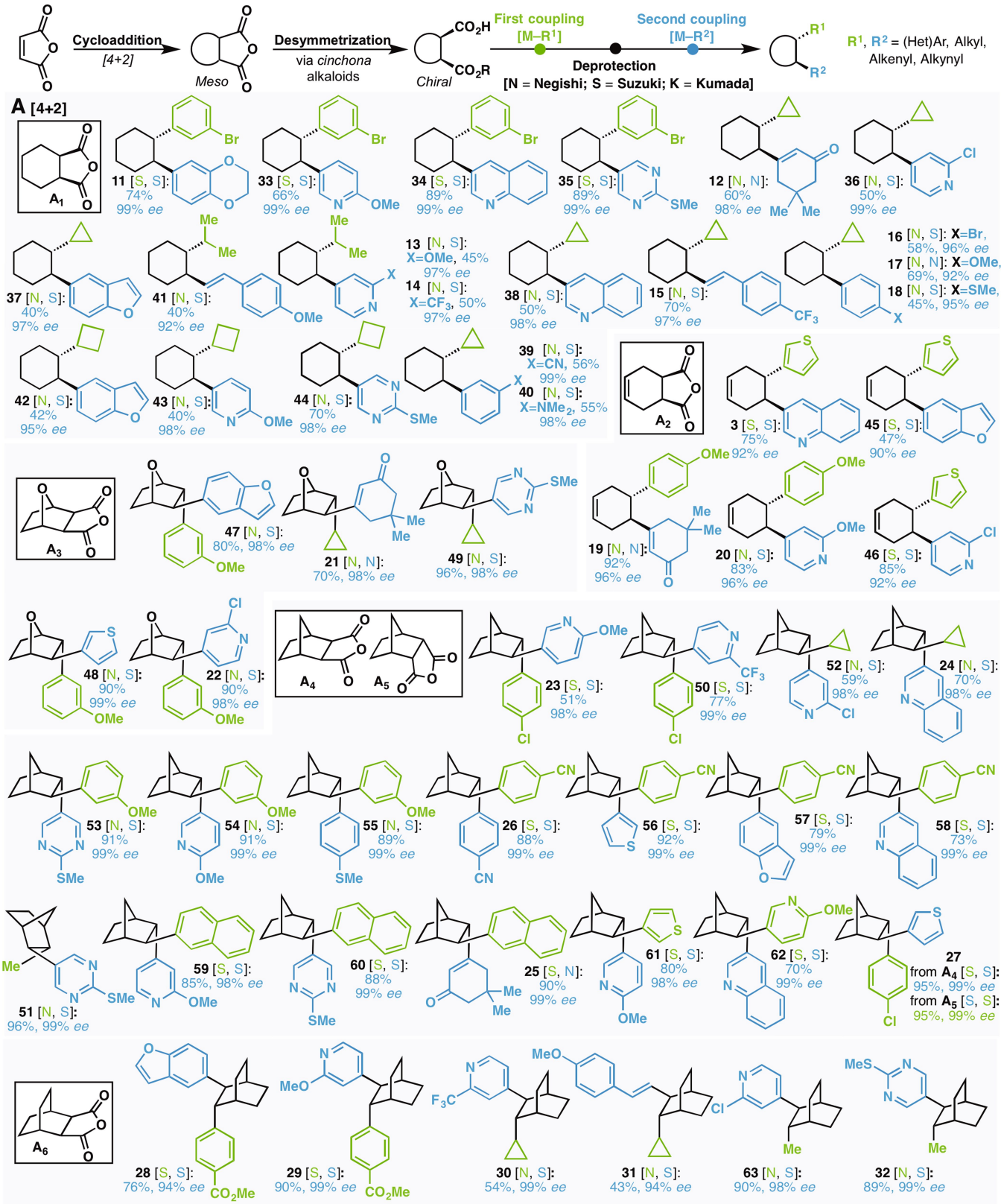
1. Fleming, I. *Pericyclic Reactions* (Oxford Univ. Press, Oxford, 2015).
2. Nicolaou, K. C., Snyder, S. A., Montagnon, T. & Vassilikogiannakis, G. The Diels–Alder reaction in total synthesis. *Angew. Chem. Int. Ed.* **41**, 1668–1698 (2002).
3. Corey, E. J. & Cheng, X. M. *The Logic of Chemical Synthesis* (Wiley, New York, 1989).
4. de Meijere, A., Bräse, S. & Oestreich, M. *Metal Catalyzed Cross-Coupling Reactions and More* (Wiley-VCH, New York, 2014).
5. Lovering, F., Bikker, J. & Humblet, C. Escape from Flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52**, 6752–6756 (2009).
6. Brown, D. G. & Boström, J. Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? *J. Med. Chem.* **59**, 4443–4458 (2016).
7. Schneider, N., Lowe, D. M., Sayle, R. A., Tarselli, M. A. & Landrum, G. A. Big data from pharmaceutical patents: a computational analysis of medicinal chemists' bread and butter. *J. Med. Chem.* **59**, 4385–4402 (2016).
8. Fleming, I. *Frontier Orbitals and Organic Chemical Reactions* (Wiley, New York, 1991).
9. Olivo, H. F. & Hemenway, M. S. Recent syntheses of epibatidine. A review. *Org. Prep. Proced. Int.* **34**, 1–25 (2002).
10. Carini, D. J. et al. Nonpeptide angiotensin II receptor antagonists: the discovery of a series of *N*-(biphenylylmethyl)imidazoles as potent, orally active antihypertensives. *J. Med. Chem.* **34**, 2525–2547 (1991).
11. Dolitzky, B.-Z., Nisnevich, G., Ruchman, I. & Kaftanov, J. Processes for preparing losartan and losartan potassium. Canadian patent CA2482857A1 (2003).
12. Larsen, R. D. et al. Efficient synthesis of losartan, a nonpeptide angiotensin II receptor antagonist. *J. Org. Chem.* **59**, 6391–6394 (1994).
13. Cornella, J. et al. Practical Ni-catalyzed aryl–alkyl cross-coupling of secondary redox-active esters. *J. Am. Chem. Soc.* **138**, 2174–2177 (2016).
14. Wang, J. et al. Nickel-catalyzed cross-coupling of redox-active esters with boronic acids. *Angew. Chem. Int. Ed.* **55**, 9676–9679 (2016).
15. Edwards, J. T. et al. Decarboxylative alkenylation. *Nature* **545**, 213–218 (2017).
16. Smith, J. M. et al. Decarboxylative alkynylation. *Angew. Chem. Int. Ed.* **56**, 11906–11910 (2017).
17. Qin, T. et al. A general alkyl-alkyl cross-coupling enabled by redox-active esters and alkylzinc reagents. *Science* **352**, 801–805 (2016).
18. Chen, Y., Tian, S.-K. & Deng, L. A highly enantioselective catalytic desymmetrization of cyclic anhydrides with modified cinchona alkaloids. *J. Am. Chem. Soc.* **122**, 9542–9543 (2000).
19. Padwa, A. & Dent, W. Use of *N*-[(trimethylsilyl)methyl]amino ethers as capped azomethine ylide equivalents. *J. Org. Chem.* **52**, 235–244 (1987).
20. Vitaku, E., Smith, D. T. & Njardarson, J. T. Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J. Med. Chem.* **57**, 10257–10274 (2014).
21. Trost, B. M. & Chan, D. M. T. Palladium-mediated cycloaddition approach to cyclopentanoids. Introduction and initial studies. *J. Am. Chem. Soc.* **105**, 2315–2325 (1983).
22. Poplata, S., Tröster, A., Zou, Y.-Q. & Bach, T. Recent advances in the synthesis of cyclobutanes by olefin [2 + 2] photocycloaddition reactions. *Chem. Rev.* **116**, 9748–9815 (2016).
23. Fan, Y.-Y., Gao, X.-H. & Yue, J.-M. Attractive natural products with strained cyclopropane and/or cyclobutane ring systems. *Sci. China Chem.* **59**, 1126–1141 (2016).
24. Bartoli, G., Bencivenni, G. & Dalpozzo, R. Asymmetric cyclopropanation reactions. *Synthesis* **46**, 979–1029 (2014).
25. Committee for Medicinal Products for Human Use (CHMP) assessment report EMA/CHMP/583011/2010 (2010).
26. Anugu, R. R., Mainkar, P. S., Sridhar, B. & Chandrasekhar, S. The Ireland-Claisen rearrangement strategy towards the synthesis of the schizophrenia drug, (+)-asenapine. *Org. Biomol. Chem.* **14**, 1332–1337 (2016).
27. Trivedi, M., Budihardjo, I., Loureiro, K., Reid, T. R. & Ma, J. D. Epothilones: a novel class of microtubule-stabilizing drugs for the treatment of cancer. *Future Oncol.* **4**, 483–500 (2008).
28. Nicolaou, K. C. et al. Chemical synthesis and biological evaluation of *cis*- and *trans*-12,13-cyclopropyl and 12,13-cyclobutyl epothilones and related pyridine side chain analogues. *J. Am. Chem. Soc.* **123**, 9313–9323 (2001).
29. Myers, M. R. et al. Potent quinoxaline-based inhibitors of PDGF receptor tyrosine kinase activity. Part 1: SAR exploration and effective bioisosteric replacement of a phenyl substituent. *Bioorg. Med. Chem. Lett.* **13**, 3091–3095 (2003).
30. Hatakeyama, T. et al. Iron-catalyzed Suzuki–Miyaura coupling of alkyl halides. *J. Am. Chem. Soc.* **132**, 10674–10676 (2010).
31. Corey, E. J., Shibata, T. & Lee, T. W. Asymmetric Diels–Alder reactions catalyzed by a triflic acid activated chiral oxazaborolidine. *J. Am. Chem. Soc.* **124**, 3808–3809 (2002).
32. He, Y. et al. The EED protein–protein interaction inhibitor A-395 inactivates the PRC2 complex. *Nat. Chem. Biol.* **13**, 389–395 (2017); erratum **13**, 922 (2017).

Acknowledgements Financial support for this work was provided by Leo Pharma and the US National Institutes of Health (NIH)/National Institute of General Medical Sciences (NIGMS; grant GM-118176). Shenzhen Haiwei M&E Co. Ltd supported a fellowship to T.–G.C.; the Uehara Memorial Foundation supported a research fellowship to S.A.; the Basque Government supported a fellowship to I.B.; Nankai University supported Y.L. and C.B.; the University of Science and Technology of China supported J.T.; and the Swiss National Science Foundation supported an Early Postdoc Mobility Fellowship to D.K. We thank L. Buzzetti for the synthesis of intermediates; D.-H. Huang and L. Pasternack for assistance with nuclear magnetic resonance spectroscopy; and A.L. Rheingold, M. Gembicky and C.E. Moore for X-ray crystallographic analysis.

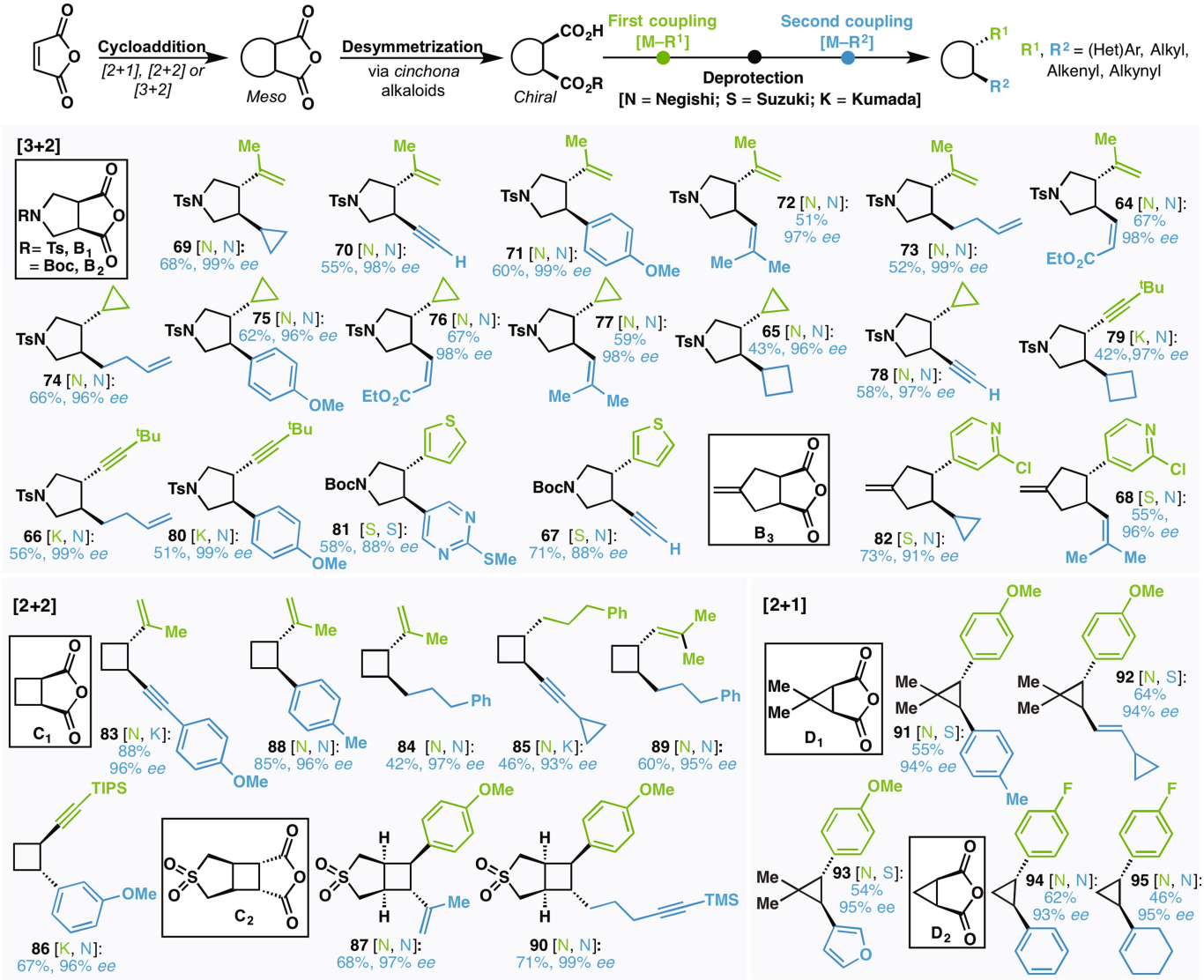
Author contributions T.–G.C., T.Q. and P.S.B. conceived the work. T.–G.C., L.M.B., Y.L., J.T., D.K., I.B., S.A., C.B., J.S.C., M.S., H.F., F.G.F., H.-W.C., L.H., T.Q. and P.S.B. designed the experiments and analysed the data. T.–G.C., L.M.B., Y.L., J.T., D.K., I.B., S.A. and C.B. performed the experiments. M.S., H.F., F.G.F., H.-W.C. and L.H. performed the experiments described in Fig. 3f. P.S.B. wrote the manuscript. T.–G.C., L.M.B., Y.L., J.T., D.K., I.B., S.A., C.B., J.S.C. and T.Q. assisted in writing and editing the manuscript.

Competing interests M.S., H.F., F.G.F., H.-W.C. and L.H. are employees of Eisai Inc. This work was part-funded by Leo Pharma.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0391-9.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0391-9.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to P.S.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Extended Data Fig. 1 | Complete substrate scope of [4+2] cycloadditions/cross-couplings. See Supplementary Information for synthetic details. R^1, R^2 = (Het)Aryl, alkyl, alkenyl, alkynyl. X-ray structure data are available for compounds **11**, **19**, **23**, **25**, **28**, **44**, **45** and **49**.



Extended Data Fig. 2 | Complete substrate scope of [3+2], [2+2] and [2+1] sections. See Supplementary Information for synthetic details. R^1, R^2 = (Het)Aryl, alkyl, alkenyl, alkynyl. X-ray structure data are available for compounds **C_2** and **65**.

LETTER

https://doi.org/10.1038/s41586-018-0413-7

Cooperative asymmetric reactions combining photocatalysis and enzymatic catalysis

Zachary C. Litman[1,7], Yajie Wang[2,7], Huimin Zhao[2,3,4,5,6]* & John F. Hartwig[1]*

Living organisms rely on simultaneous reactions catalysed by mutually compatible and selective enzymes to synthesize complex natural products and other metabolites. To combine the advantages of these biological systems with the reactivity of artificial chemical catalysts, chemists have devised sequential, concurrent, and cooperative chemoenzymatic reactions that combine enzymatic and artificial catalysts[1–9]. Cooperative chemoenzymatic reactions consist of interconnected processes that generate products in yields and selectivities that cannot be obtained when the two reactions are carried out sequentially with their respective substrates[2,7]. However, such reactions are difficult to develop because chemical and enzymatic catalysts generally operate in different media at different temperatures and can deactivate each other[1–9]. Owing to these constraints, the vast majority of cooperative chemoenzymatic processes that have been reported over the past 30 years can be divided into just two categories: chemoenzymatic dynamic kinetic resolutions of racemic alcohols and amines, and enzymatic reactions requiring the simultaneous regeneration of a cofactor[2,4,5]. New approaches to the development of chemoenzymatic reactions are needed to enable valuable chemical transformations beyond this scope. Here we report a class of cooperative chemoenzymatic reaction that combines photocatalysts that isomerize alkenes with ene-reductases that reduce carbon–carbon double bonds to generate valuable enantioenriched products. This method enables the stereoconvergent reduction of *E*/*Z* mixtures of alkenes or reduction of the unreactive stereoisomers of alkenes in yields and enantiomeric excesses that match those obtained from the reduction of the pure, more reactive isomers. The system affords a range of enantioenriched precursors to biologically active compounds. More generally, these results show that the compatibility between photocatalysts and enzymes enables chemoenzymatic processes beyond cofactor regeneration and provides a general strategy for converting stereoselective enzymatic reactions into stereoconvergent ones.

To develop a cooperative chemoenzymatic reaction (Fig. 1) that enables a stereoconvergent enzymatic reduction of an isomeric mixture of alkenes, initial reactions were conducted with the model substrate 2-phenylbut-2-enedioic acid dimethyl ester (**1a**). YersER, an ene-reductase isolated from *Yersinia bercovieri*, exclusively reduces the *E* isomer of **1a** ((*E*)-**1a**) to dimethyl 2-phenylsuccinate (**2a**) in high yield and with excellent enantioselectivity in the presence of a glucose dehydrogenase enzyme for cofactor regeneration (Supplementary Table 3). The observed selectivity results from interactions between the *E* and *Z* isomers of **1a** and the active site of the enzyme[10–13]. These

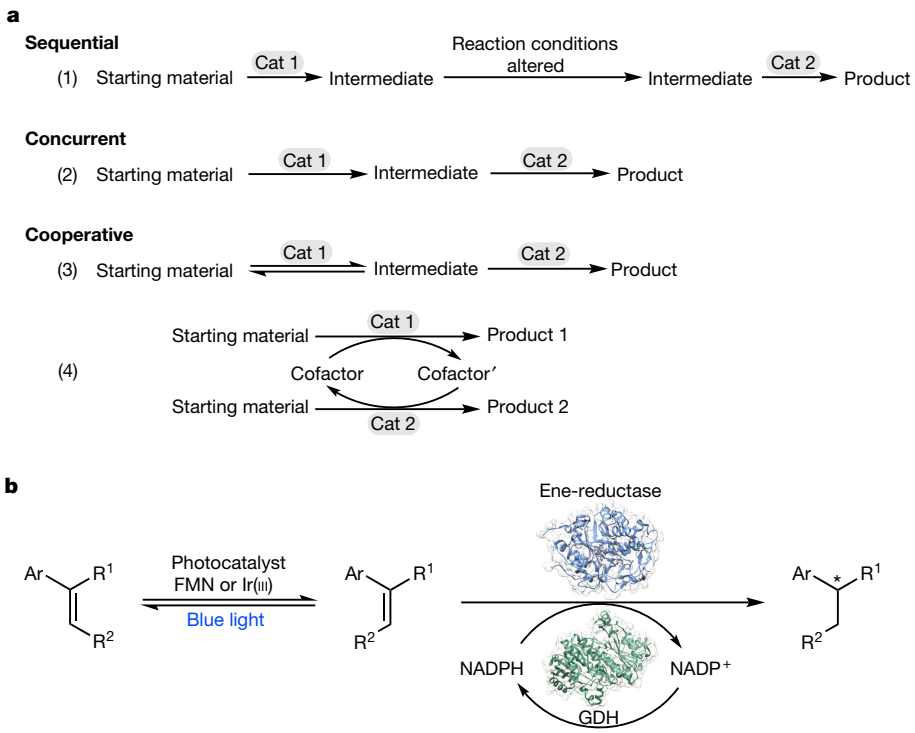
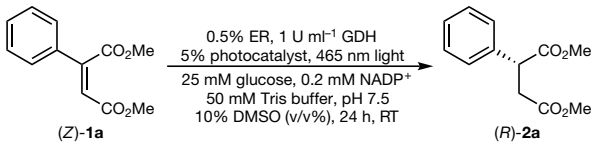


Fig. 1 | Chemoenzymatic reactions. a, Types of chemoenzymatic reaction. At least one enzymatic catalyst and one chemical catalyst is present in all reactions. (1) Reactions conducted in a sequential manner with an intermediate step in which reaction conditions are altered or reagents or catalysts are added. (2) Two irreversible reactions conducted in a simultaneous manner. (3) A cooperative reaction consisting of a reversible reaction combined with an irreversible reaction run simultaneously. (4) A cooperative reaction in which a sub-stoichiometric quantity of cofactor or reagent is recycled and both catalysts operate simultaneously. The prime indicates an altered oxidation state of the cofactor. See Supplementary Fig. 1 for additional explanation and definitions. **b**, Combination of photocatalytic isomerization and enzymatic reduction of alkenes. The asterisk indicates the chiral center. Ar, aryl; GDH, glucose dehydrogenase; NADP(H), (reduced) nicotinamide adenine dinucleotide phosphate.

[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [4]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [5]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [6]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [7]These authors contributed equally: Zachary C. Litman, Yajie Wang. *e-mail: zhao5@illinois.edu; jhartwig@berkeley.edu

a



Entry	Photocatalyst	PS (% *E*-**1a**)	Yield (%)
1	None*	0	0
2	None	3	1
3	Ribo	90	18
4	FMN	90	70
5	FAD	90	54
6	EosY	69	21
7	Fluor	46	4
8	RuBpy	68	2
9	RuBpz	65	2
10	**Ir-16**	79	83
11	**Ir-67**	82	70
12	**Ir-80**	74	64

b

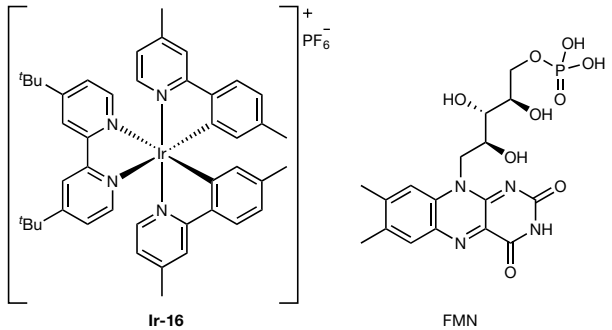


Fig. 2 | Simultaneous isomerization and reduction of (*Z*)-1a. **a**, Reduction of (*Z*)-**1a** in the presence of a series of photocatalysts and blue light. Photostationary state (PS) is the percentage of (*E*)-**1a** present after (*Z*)-**1a** was irradiated with blue light in the presence of 5% photocatalyst for 24 h. Experiments labelled 'None*' were conducted in the absence of a photocatalyst, but the presence of ambient light; those labelled 'None' were conducted in the absence of a photocatalyst, but the presence of blue light. The photostationary state values for experiments labelled 'None*' and 'None' have been corrected for initial trace quantities of (*E*)-**1a**. **b**, Structures of **Ir-16** and FMN. EosY, eosin Y; ER, ene-reductase; FAD, flavin adenine dinucleotide; Fluor, fluorescein; FMN, flavin mononucleotide; Ribo, riboflavin; RT, room temperature; RuBpy, $[Ru(bpy)_3]Cl_2$; RuBpz, $[Ru(bpz)_3](PF_6)_2$ (bpy, 2,2′-bipyridine; bpz, 2,2′-bipyrazine).

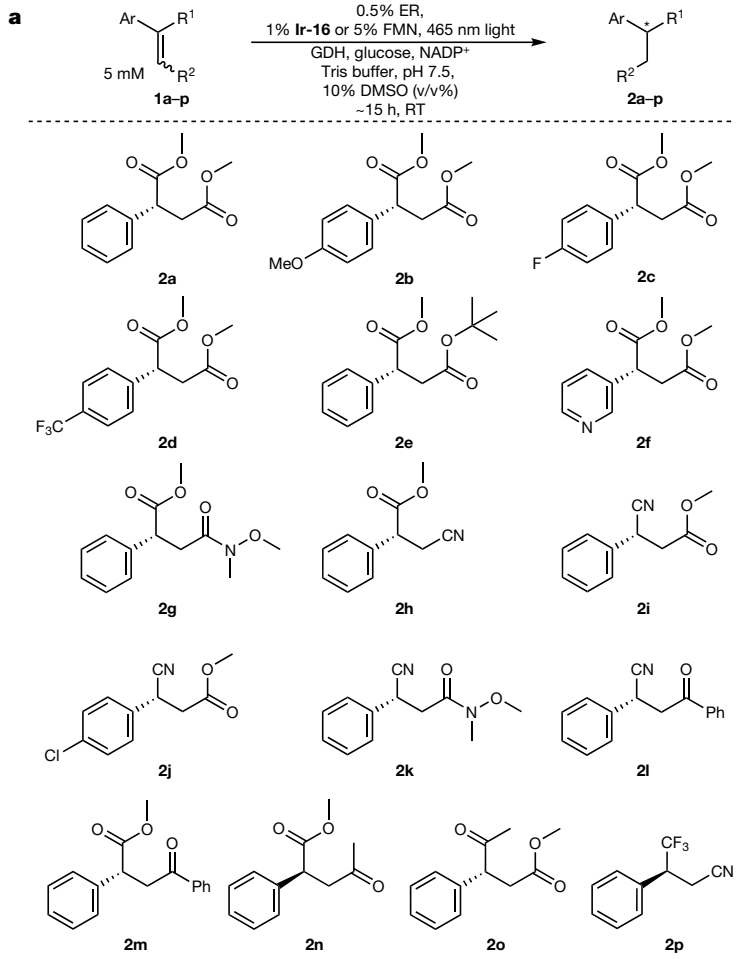
interactions have been investigated for related diester and cyanoacrylate substrates[14–16]. To enable a stereoconvergent reduction reaction, it was necessary to identify a catalyst for the *E*/*Z* isomerization of olefins that is compatible with ene-reductases. An appropriate isomerization catalyst would operate in aqueous solution at ambient temperature; remain active in the presence of ene-reductases, substrates, and products in the reaction mixture; isomerize olefins at the low substrate concentrations required for enzymatic reduction; and generate the more reactive isomer of a substrate from the less reactive isomer. The isomerization catalyst must also be tolerated by the ene-reductases; be mutually compatible with a regeneration system comprising glucose dehydrogenase and reduced nicotinamide adenine dinucleotide phosphate; and not racemize the product.

After determining that negligible thermal isomerization of (*Z*)-**1a** occurs under ambient conditions (Fig. 2a), we considered that recently reported photocatalytic isomerizations of alkenes[17–21] could be combined with enzymatic reduction to develop the proposed cooperative chemoenzymatic process. Our initial experiments assessed the photoisomerization of (*Z*)-**1a**, in solvent mixtures appropriate for the reduction of alkenes by ene-reductases. Limited isomerization of (*Z*)-**1a** was observed in semi-aqueous media (1:9 DMSO:Tris buffer) in the presence of blue light (450–470 nm) and in the absence of a photocatalyst (Fig. 2a). However, extensive isomerization of (*Z*)-**1a** was observed when the reaction was conducted with riboflavin in the presence of blue light. A photostationary state consisting of a 9:1 ratio of (*E*)-**1a** to (*Z*)-**1a** was obtained after (*Z*)-**1a** was irradiated with blue light for 24 h in the presence of 5% riboflavin (Fig. 2a). However, modest yields of **2a** were obtained when YersER and riboflavin were used in a simultaneous isomerization and reduction of (*Z*)-**1a** in the presence of blue light. This result demonstrated the importance of establishing compatibility between the enzyme and the photocatalyst. We proposed that competitive binding of riboflavin to the flavin-binding site of YersER led to the inhibition of enzymatic activity[13]. Therefore, we sought alternative photocatalysts that would lead to isomerization of (*Z*)-**1a** without inhibiting YersER or the cofactor-regeneration system.

The *E*/*Z* isomerization of (*Z*)-**1a** was evaluated in the presence of a series of organometallic and organic photocatalysts in 1:9 DMSO:50 mM Tris buffer, and the photostationary state for each combination is recorded in Fig. 2. Greater than 40% conversion of (*Z*)-**1a** to (*E*)-**1a** was observed when the substrate was irradiated in the presence of the majority of the photocatalysts, and the highest *E*/*Z* ratios exceeded 8:1. Increasing catalyst loading and increasing light intensity enhanced the rate of photoisomerization of (*Z*)-**1a** (Supplementary Figs. 15, 16, 18).

Having identified a series of photocatalysts for the isomerization of (*Z*)-**1a** to (*E*)-**1a** in a semi-aqueous medium, we evaluated the simultaneous, cooperative photoisomerization and enzymatic reduction of (*Z*)-**1a** with the same photocatalysts (Fig. 2a). Moderate to high yields of **2a** were obtained when a range of catalysts were used in the cooperative process. The highest conversions and yields were obtained when the cooperative reactions were conducted with 5% of flavin mononucleotide (FMN) or 5% of the cationic iridium (III) complexes $[Ir(dmppy)_2(dtbbpy)]PF_6$ (**Ir-16**), $[Ir(dtbbpy)(ppy)_2]PF_6$ (**Ir-67**) and $[Ir(dtbppy)_2(dtbbpy)]PF_6$ (**Ir-80**) (dmppy, 4-methyl-2-(4-methylphenyl)pyridine; dtbbpy, 4,4′-di-*tert*-butyl-2,2′-bipyridine; ppy, 2-phenylpyridine; dtbppy, 4-(*tert*-butyl)-2-(4-(*tert*-butyl)phenyl)pyridine.)

The high yields and enantioselectivities obtained from the cooperative reduction of the model diester (*Z*)-**1a** encouraged us to investigate the cooperative reduction of other aryl diesters. We identified enzymes that preferentially reduce the *E* isomers of diesters **1b–f** (Fig. 3a) in high yields and enantioselectivities (Supplementary Tables 4–8). These enzymes were then used in the cooperative reduction of the *Z* isomers of **1b–1d** with **Ir-16** or FMN as a photocatalyst in the presence of blue light. High yields of product were obtained from the cooperative isomerization and reduction of each of the diesters, including those containing electron-rich ((*Z*)-**1b**) and electron-poor ((*Z*)-**1d**) aryl groups with either FMN or **Ir-16** as photocatalysts and the enzymes YersER or XenB (Fig. 3a). The yields and enantiomeric excess (e.e.) of the products obtained from the cooperative reductions of the *Z* isomers of **1b–d** were equivalent to those obtained from the enzymatic reduction of the *E* isomers of substrates **1b–d** in the absence of the photocatalyst. This result suggests that the cooperative reduction of any *E*/*Z* mixture of these alkenes should give high yields of the reduced products and proceed with high enantioselectivity. This feature of the cooperative system was crucial to obtain high yields and enantioselectivities for the reduction of **1e** and **1f**, which were synthesized as inseparable mixtures of *E* and *Z* alkenes. The cooperative chemoenzymatic reduction of a 62:38 mixture of *E* and *Z* isomers of **1e** afforded **2e** in 74% yield and >99% e.e., and the cooperative reduction of a 61:39 mixture of *E* and *Z* isomers of **1f** afforded **2f** in 94% yield and 91% e.e. The yields and enantioselectivities from the cooperative reduction of (*E*/*Z*)-**1e** and (*E*/*Z*)-**1f** indicate that the reactions were stereoconvergent. The enzymatic reductions of **1e** and **1f** in the absence of a photocatalyst and light formed the reduced products in only 58% and 60%



Substrate	Enzyme	Photocatalyst	PS	Yield (%)	e.e. (%)
(*Z*)-**1a**	YersER	**Ir-16**	78% *E*	87	>99
		FMN	94% *E*	70	>99
(*Z*)-**1b**	YersER	**Ir-16**	69% *E*	89	>99
		FMN	61% *E*	63	92
(*Z*)-**1c**	XenB	**Ir-16**	68% *E*	89	>99
(*Z*)-**1d**	XenB	**Ir-16**	72% *E*	82	>99
(62:38 *E*/*Z*)-**1e**	OPR1	**Ir-16**	67% *E*	74	>99
(61:39 *E*/*Z*)-**1f**	XenB	**Ir-16**	81% *E*	94	91
(*Z*)-**1g**	OPR1	**Ir-16**	88% *E*	82	>99
(*Z*)-**1h**	TOYE[a]	**Ir-16**	68% *E*	87	>99
		FMN[b]	69% *E*	98	>99
(*Z*)-**1i**	OYE2[a]	**Ir-16**	61% *E*	92	98
		FMN[b]	46% *E*	94	99
(*Z*)-**1j**	OYE2	**Ir-16**[c]	9% *E*	80	92
		FMN	3% *E*	60	90
(*Z*)-**1k**	YersER[a]	**Ir-16**	70% *E*	96	99
		FMN	71% *E*	82	99
(*Z*)-**1l**	OYE2[a]	–[d]	89% *E*	95	>99
(*Z*)-**1m**	OPR1	**Ir-16**	83% *E*	81	>99
		–[d]	92% *E*	80	>99
(*Z*)-**1n**	SYE1	**Ir-16**	78% *E*	60	88
(*Z*)-**1o**	YersER	**Ir-16**	76% *E*	92	96
		FMN	89% *E*	90	86
(*E*)-**1p**	OYE2	**Ir-16**	14% *Z*	74	>99

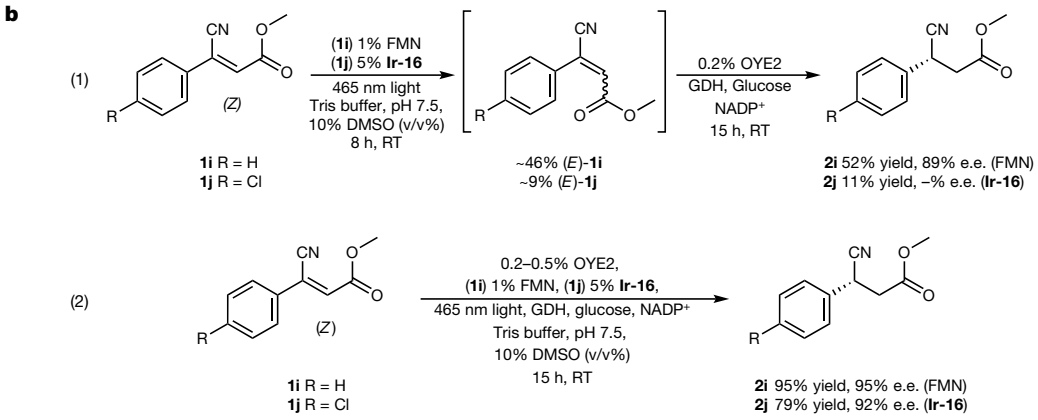


Fig. 3 | Scope of the cooperative photoisomerization and reduction and comparison to sequential reactions. a, Scope of the cooperative reduction of alkenes. Standard conditions: reactions were conducted with 0.5% ene-reductase, 1 U ml^{-1} glucose dehydrogenase, 1% **Ir-16** or 5% FMN, 25 mM glucose, 0.2 mM $NADP^+$, 50 mM Tris buffer at pH 7.5 and 10% DMSO at room temperature in the presence of blue light. PS is the percentage of *E* substrate present after the *Z* isomer (or *E*/*Z* isomer mixture) was irradiated with light in the presence of either 5% FMN or 1% **Ir-16** for 24 h. In the case of **1p**, (*E*)-**1p** was irradiated with blue light in the presence of FMN and **Ir-16** for 24 h, and the percentage of (*Z*)-**1p** was determined. Deviations from standard conditions are as follows: [a]0.2% ene-reductase; [b]1% FMN; [c]5% **Ir-16**; [d]the reaction was conducted in the absence of a photocatalyst in 5% DMSO, 95% 50 mM Tris buffer. Control experiments are included in Supplementary Tables 3–18. **b**, Comparison of sequential (1) and cooperative (simultaneous) (2) isomerization and reduction of cyanoacrylates **1i** and **1j** under standard conditions.

yield, which reflects reaction with only the *E* isomer of the *E*/*Z* mixture (Supplementary Tables 7, 8).

To determine if the cooperative reaction would give high yields and high enantioselectivities with alkenes other than diesters, the cooperative reductions of unsaturated compounds containing diverse combinations of functional groups were evaluated. The ene-reductases OPR1, TOYE, OYE2, YersER, and SYE1 preferentially reduced the *E* isomers of alkenes **1g**–**o** to form **2g**–**o** in high yields and with high enantioselectivities (Supplementary Tables 9–17). The cooperative reactions of the *Z* isomers of **1g**–**o** were then conducted with **Ir-16** or FMN in the presence of blue light. Figure 3a shows results from the combination of a photocatalyst and an enzyme that generated the products in the highest yields and with the highest enantioselectivities from the *Z* isomers of **1g**–**o**. Good results were obtained for the cooperative reduction of

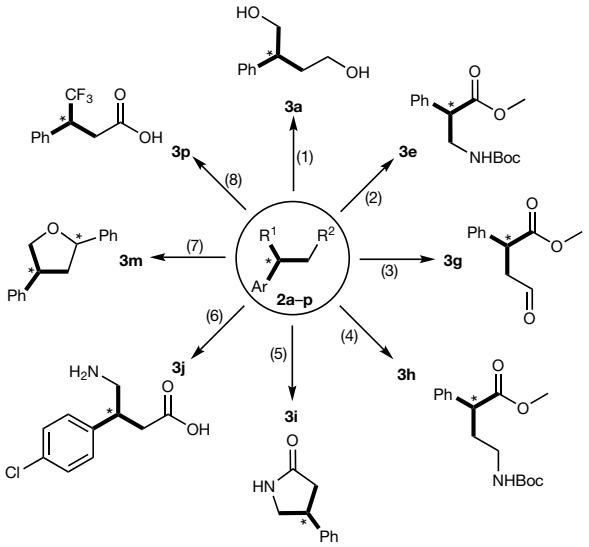


Fig. 4 | Derivatization of enantioenriched products. A summary of the previously reported and newly disclosed transformations of the products of the chemoenzymatic cooperative reduction of alkenes is shown. Reaction conditions for derivatizations are as follows: (1) **2a** to **3a**: 4.0 equiv. $LiAlH_4$, tetrahydrofuran, RT, 5 h. (2) **2e** to **3e**: i, 1:1 trifluoroacetic acid:dichloromethane, 3 h, RT. ii, 1.7 equiv. triethylamine, dichloroethane, and 1.5 equiv. diphenylphosphoryl azide, 65 °C, 2 h. iii, 0.10 equiv. $Mo(O)_2Cl_2$, *tert*-butanol, 65 °C, 1 h 10 min. (3) **2g** to **3g**: 1.4 equiv. $Cp_2Zr(H)Cl$ (Cp, cyclopentadiene), 1.0 equiv. ZrF_4, 1:1 tetrahydrofuran:dichloromethane, 3 h, RT. (4) **2h** to **3h** was previously reported[27]. (5) **2i** to **3i** was previously reported[14]. (6) **2j** to **3j** was previously reported[14]. (7) **2m** to **3m** was previously reported[28]. (8) **2p** to **3p**: 1:1 H_2O:H_2SO_4, 120 °C, 6 h.

β-cyano-α,β-unsaturated ester (*Z*)-**1h**, α-cyano-α,β-unsaturated esters (*Z*)-**1i** and (*Z*)-**1j**, amidoacrylate (*Z*)-**1g**, and amidocyanate (*Z*)-**1k**. High yields and enantioselectivities were also obtained from the reactions of cyanoketone (*Z*)-**1l**, β-keto-α,β-unsaturated esters (*Z*)-**1m** and (*Z*)-**1n**, and α-keto-α,β-unsaturated ester (*Z*)-**1o**. The cooperative reductions of amidoacrylate (*Z*)-**1g** and amidocyanate (*Z*)-**1k** are noteworthy because enzymatic reductions of alkenes containing Weinbreb amides have not been reported to our knowledge. Control experiments showed that (*Z*)-**1l**, (*Z*)-**1m**, and (*Z*)-**1n** undergo enzymatic reduction to generate products in high yields and with high enantioselectivities in the presence of blue light but in the absence of an added photocatalyst (Supplementary Tables 14–16). In these cases, the alkenes (**1l–n**) isomerize in the presence of blue light alone (Supplementary Figs. 13, 22, 23).

The cooperative enzymatic reductions of alkenes **1i** and **1j** shown in Fig. 3b illustrate the benefits of a cooperative system over two sequential reactions. The photoisomerization of (*Z*)-**1i** and (*Z*)-**1j** with **Ir-16** or FMN results in *E*/*Z* mixtures in which the less reactive *Z* isomer is the major component. As a result, low yields were obtained from the sequential isomerization and reduction of (*Z*)-**1i** and (*Z*)-**1j**. The e.e. of **2i** obtained from the sequential isomerization and reduction of (*Z*)-**1i** was slightly lower than the e.e. obtained from the enzymatic reduction of the pure, more reactive isomer (*E*)-**1i**, and this difference probably results from the slow reduction of the *Z* isomer of **1i** after rapid consumption of the *E* isomer during the second stage of the sequential process (Supplementary Fig. 30). By contrast, the simultaneous, cooperative reduction of (*Z*)-**1i** and (*Z*)-**1j** generated products **2i** and **2j** in high yields and with high enantioselectivities.

An enzyme that selectively reduces the *E* isomer of trifluoromethylcyanate **1p** in high yield and with high enantioselectivity could not be identified; however, OYE2 selectively reduces the *Z* isomer of **1p** in this way (Supplementary Table 18). To determine if a cooperative reaction could convert (*E*)-**1p** to **2p** in high yield and with high enantioselectivity, the ability of (*E*)-**1p** to undergo photoisomerization in the presence of **Ir-16** and blue light was evaluated. An 86:14 ratio of the *E*/*Z* isomers of **1p** was established after 24 h of irradiation with blue light in the presence of **Ir-16**. Although the photostationary state of **1p** favours the less reactive *E* isomer, the cooperative reaction of (*E*)-**1p** to **2p** occurred in high yield and with high enantioselectivity with **Ir-16** as the photocatalyst and OYE2 as the reductase. This example illustrates an important benefit of the cooperative chemoenzymatic reaction: the system enables either the isomerization of a *Z* alkene with simultaneous enzymatic reduction of the *E* isomer or the isomerization of an *E* alkene with simultaneous enzymatic reduction of the *Z* isomer. The conversion of (*E*)-**1p** to **2p** is also noteworthy because of the dearth of reductions of an alkene with isolated enzymes to generate a product containing a stereogenic centre substituted with a trifluoromethyl group.

To demonstrate the synthetic value of this new method, preparative-scale cooperative reactions were conducted with 1% **Ir-16** and the following alkenes: 40–60 mg of (*Z*)-**1a**, a 62:38 mixture of the *E* and *Z* isomers of **1e**, (*Z*)-**1g**, (*Z*)-**1h**, and (*Z*)-**1o**. Product **2a** was isolated in 87% yield and >99% e.e., **2e** in 79% yield and >99% e.e., **2g** in 71% yield and >99% e.e., **2h** in 96% yield and 92% e.e., and **2o** in 79% yield and >99% e.e.

The enantioenriched compounds that were obtained from the cooperative isomerization and enzymatic reduction system can be transformed into a variety of biologically active molecules and valuable synthetic intermediates (Fig. 4). For example, the selective hydrolysis of the *tert*-butyl ester in compound **2e** followed by a Curtius rearrangement yielded **3e**, a β^2-amino ester. The β-amino ester was isolated in 90% yield without a notable reduction in enantiomeric excess (98% e.e.). The selective reduction of the Weinreb amide in compound **2g** with Schwartz's reagent yielded methyl 4-oxo-2-phenylbutanoate (**3g**) in 74% isolated yield and 99% e.e. This compound is an intermediate in the synthesis of protein kinase inhibitors and microsomal triglyceride transfer protein inhibitors[22,23]. Acid-catalysed hydrolysis of the nitrile in **2p** yielded 4,4,4-trifluoro-3-phenylbutanoic acid (**3p**) in 96% yield and >99% e.e. This versatile synthetic intermediate has previously been used in the synthesis of inhibitors of beta amyloid production[24,25]. Reduction of **2a** with lithium aluminium hydride formed 2-phenylbutane-1,4-diol (**3a**) in 93% yield and >99% e.e. This diol is a synthetic precursor to inhibitors of matrix metalloproteases[26]. Other products of the cooperative reactions are known precursors to biologically active compounds. For example, **2i** and **2j** have been converted previously to γ-amino acids (including baclofen and phenibut) and γ-lactams[14], **2h** has been converted into a γ^2 amino ester and a γ^2 lactam[27], and **2m** has been converted into calyxolanes, which are cyclic ether natural products[28]. Thus, the products of the cooperative chemoenzymatic reduction are precursors to valuable synthetic intermediates using both newly disclosed and previously reported transformations.

The combination of a photocatalytic process and an enzymatic reaction enables transformations that combine the reactivity of chemical catalysts with the selectivity of enzymes. Two features of photocatalytic reactions make them, in general, suitable for chemoenzymatic processes: first, photochemical reactions typically occur at or near room temperature, making them compatible with the thermal requirements of enzymatic systems, and second, photocatalysts often react by mechanisms, such as outer-sphere electron transfer or energy transfer, that involve intermediates that are stable towards water and the functional groups in proteins. These considerations, in combination with the renewed interest in photocatalysis and the rapidly advancing tools of molecular biology, should create opportunities for the development of a wide range of new cooperative chemoenzymatic transformations.

Data availability

Data supporting the findings of this study, Supplementary Figures 1–42 and Supplementary Tables 1–19 are available as Supplementary Information. Any additional information is available from the corresponding author upon reasonable request.

Received: 21 March; Accepted: 25 June 2018;
Published online 15 August 2018.

1. Rudroff, F. et al. Opportunities and challenges for combining chemo- and biocatalysis. *Nat. Catal.* **1**, 12–22 (2018).
2. Köhler, V. & Turner, N. J. Artificial concurrent catalytic processes involving enzymes. *Chem. Commun.* **51**, 450–464 (2015).
3. Denard, C. A., Hartwig, J. F. & Zhao, H. Multistep one-pot reactions combining biocatalysts and chemical catalysts for asymmetric synthesis. *ACS Catal.* **3**, 2856–2864 (2013).
4. Gröger, H. in *Cooperative Catalysis: Designing Efficient Catalysts for Synthesis* (ed. Peters, R.) Ch. 11, 325–350 (Wiley, Weinheim, 2015).
5. Verho, O. & Bäckvall, J.-E. Chemoenzymatic dynamic kinetic resolution: a powerful tool for the preparation of enantiomerically pure alcohols and amines. *J. Am. Chem. Soc.* **137**, 3996–4009 (2015).
6. Wang, Y., Ren, H. & Zhao, H. Expanding the boundary of biocatalysis: design and optimization of *in vitro* tandem catalytic reactions for biochemical production. *Crit. Rev. Biochem. Mol. Biol.* **53**, 115–129 (2018).
7. Denard, C. A. et al. Cooperative tandem catalysis by an organometallic complex and a metalloenzyme. *Angew. Chem. Int. Ed.* **53**, 465–469 (2014).
8. Köhler, V. et al. Synthetic cascades are enabled by combining biocatalysts with artificial metalloenzymes. *Nat. Chem.* **5**, 93–99 (2013).
9. Haak, R. M. et al. Dynamic kinetic resolution of racemic β-haloalcohols: direct access to enantioenriched epoxides. *J. Am. Chem. Soc.* **130**, 13508–13509 (2008).
10. Yanto, Y. et al. Asymmetric bioreduction of alkenes using ene-reductases YersER and KYE1 and effects of organic solvents. *Org. Lett.* **13**, 2540–2543 (2011).
11. Stueckler, C. et al. Stereocomplementary bioreduction of α,β-unsaturated dicarboxylic acids and dimethyl esters using enoate reductases: enzyme- and substrate-based stereocontrol. *Org. Lett.* **9**, 5409–5411 (2007).
12. Gatti, F. G., Parmeggiani, F. & Sacchetti, A. in *Synthetic Methods for Biologically Active Molecules* 1st edn (ed. Brenna, E.) Ch. 3, 49–84 (Wiley, Weinheim, 2013).
13. Chaparro-Riggers, J. F. et al. Comparison of three enoate reductases and their potential use for biotransformations. *Adv. Synth. Catal.* **349**, 1521–1531 (2007).
14. Brenna, E. et al. Opposite enantioselectivity in the bioreduction of (*Z*)-β-aryl-β-cyanoacrylates mediated by the tryptophan 116 mutants of old yellow enzyme 1: synthetic approach to (*R*)- and (*S*)-β-aryl-γ-lactams. *Adv. Synth. Catal.* **357**, 1849–1860 (2015).
15. Brenna, E. et al. Old yellow enzyme-mediated reduction of β-cyano-α,β-unsaturated esters for the synthesis of chiral building blocks: Stereochemical analysis of the reaction. *Catal. Sci. Technol.* **3**, 1136–1146 (2013).
16. Wang, Y. et al. Combining Rh-catalyzed diazocoupling and enzymatic reduction to efficiently synthesize enantioenriched 2-substituted succinate derivatives. *ACS Catal.* **7**, 2548–2552 (2017).
17. Metternich, J. B. & Gilmour, R. A bio-inspired, catalytic *E* → *Z* isomerization of activated olefins. *J. Am. Chem. Soc.* **137**, 11254–11257 (2015).
18. Singh, K., Staig, S. J. & Weaver, J. D. Facile synthesis of *Z*-alkenes via uphill catalysis. *J. Am. Chem. Soc.* **136**, 5275–5278 (2014).
19. Metternich, J. B. et al. Photocatalytic *E*→*Z* isomerization of polarized alkenes inspired by the visual cycle: Mechanistic dichotomy and origin of selectivity. *J. Org. Chem.* **82**, 9955–9977 (2017).
20. Metternich, J. B. & Gilmour, R. Photocatalytic *E*→*Z* isomerization of alkenes. *Synlett* **27**, 2541–2552 (2016).
21. Molloy, J. J. et al. Contra-thermodynamic, photocatalytic *E*→*Z* isomerization of styrenyl boron species: vectors to facilitate exploration of two-dimensional chemical space. *Angew. Chem. Int. Ed.* **57**, 3168–3172 (2018).
22. Bamaung, N. Y. et al. Protein kinase inhibitors. US patent 2009/US20090253723 Al (2009).
23. Meerpoel, L. et al. Mtp inhibiting aryl piperidines or piperazines substituted with 5-membered heterocycles. US patent 2007/US20070191383 Al (2007).
24. Caggiano, T. J. et al. Inhibitors of beta amyloid production. US patent 2009/US20090023801 Al (2009).
25. Dong, K., Li, Y., Wang, Z. & Ding, K. Catalytic asymmetric hydrogenation of α-CF_3- or β-CF_3-substituted acrylic acids using rhodium(I) complexes with a combination of chiral and achiral ligands. *Angew. Chem. Int. Ed.* **52**, 14191–14195 (2013).
26. Chung, Y.-C., Janmanchi, D. & Wu, H.-L. Preparation of chiral 3-arylpyrrolidines via the enantioselective 1,4-addition of arylboronic acids to fumaric esters catalyzed by Rh(I)/chiral diene complexes. *Org. Lett.* **14**, 2766–2769 (2012).
27. Kong, D., Li, M., Wang, R., Zi, G. & Hou, G. Highly efficient asymmetric hydrogenation of cyano-substituted acrylate esters for synthesis of chiral γ-lactams and amino acids. *Org. Biomol. Chem.* **14**, 1216–1220 (2016).
28. Zhao, F. et al. Enantioselective aza-ene-type reactions of enamides with gold carbenes generated from α-diazoesters. *Angew. Chem. Int. Ed.* **56**, 3247–3251 (2017).

Acknowledgements We thank K. Faber, U. Bornscheuer and N. Scrutton for the gift of plasmids pET21a-OPR1, pGaston-XenB and pET21b_TOYE respectively. We also thank the Metabolomics Center of University of Illinois at Urbana-Champaign for gas chromatography–mass spectrometry (GC–MS) facilities and A. Vladimirovich Ulanov for suggestions on GC analysis. This work was supported by the National Science Foundation under the CCI Center for Enabling New Technologies through Catalysis (CENTC, CHE-1205189 to H.Z. and J.F.H.) and Department of Energy (DE-SC0018420 to H.Z.).

Reviewer information *Nature* thanks R. Gilmour, W. Wang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Z.C.L., Y.W., J.F.H. and H.Z. conceived the project, designed the initial experiments and interpreted the data. Z.C.L. and Y.W. performed all of the experiments. Z.C.L., Y.W., J.F.H. and H.Z. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0413-7.

Correspondence and requests for materials should be addressed to H.Z. or J.F.H.

LETTER

https://doi.org/10.1038/s41586-018-0383-9

Marine heatwaves under global warming

Thomas L. Frölicher[1,2]*, Erich M. Fischer[3] & Nicolas Gruber[4]

Marine heatwaves (MHWs) are periods of extreme warm sea surface temperature that persist for days to months[1] and can extend up to thousands of kilometres[2]. Some of the recently observed marine heatwaves revealed the high vulnerability of marine ecosystems[3–11] and fisheries[12–14] to such extreme climate events. Yet our knowledge about past occurrences[15] and the future progression of MHWs is very limited. Here we use satellite observations and a suite of Earth system model simulations to show that MHWs have already become longer-lasting and more frequent, extensive and intense in the past few decades, and that this trend will accelerate under further global warming. Between 1982 and 2016, we detect a doubling in the number of MHW days, and this number is projected to further increase on average by a factor of 16 for global warming of 1.5 degrees Celsius relative to preindustrial levels and by a factor of 23 for global warming of 2.0 degrees Celsius. However, current national policies for the reduction of global carbon emissions are predicted to result in global warming of about 3.5 degrees Celsius by the end of the twenty-first century[16], for which models project an average increase in the probability of MHWs by a factor of 41. At this level of warming, MHWs have an average spatial extent that is 21 times bigger than in preindustrial times, last on average 112 days and reach maximum sea surface temperature anomaly intensities of 2.5 degrees Celsius. The largest changes are projected to occur in the western tropical Pacific and Arctic oceans. Today, 87 per cent of MHWs are attributable to human-induced warming, with this ratio increasing to nearly 100 per cent under any global warming scenario exceeding 2 degrees Celsius. Our results suggest that MHWs will become very frequent and extreme under global warming, probably pushing marine organisms and ecosystems to the limits of their resilience and even beyond, which could cause irreversible changes.

There is mounting evidence that global warming is leading to more frequent and intense heatwaves over land, increasing the risk of severe and in some cases irreversible impacts[17]. In comparison, we know much less about how heatwaves in the ocean unfold in time and what

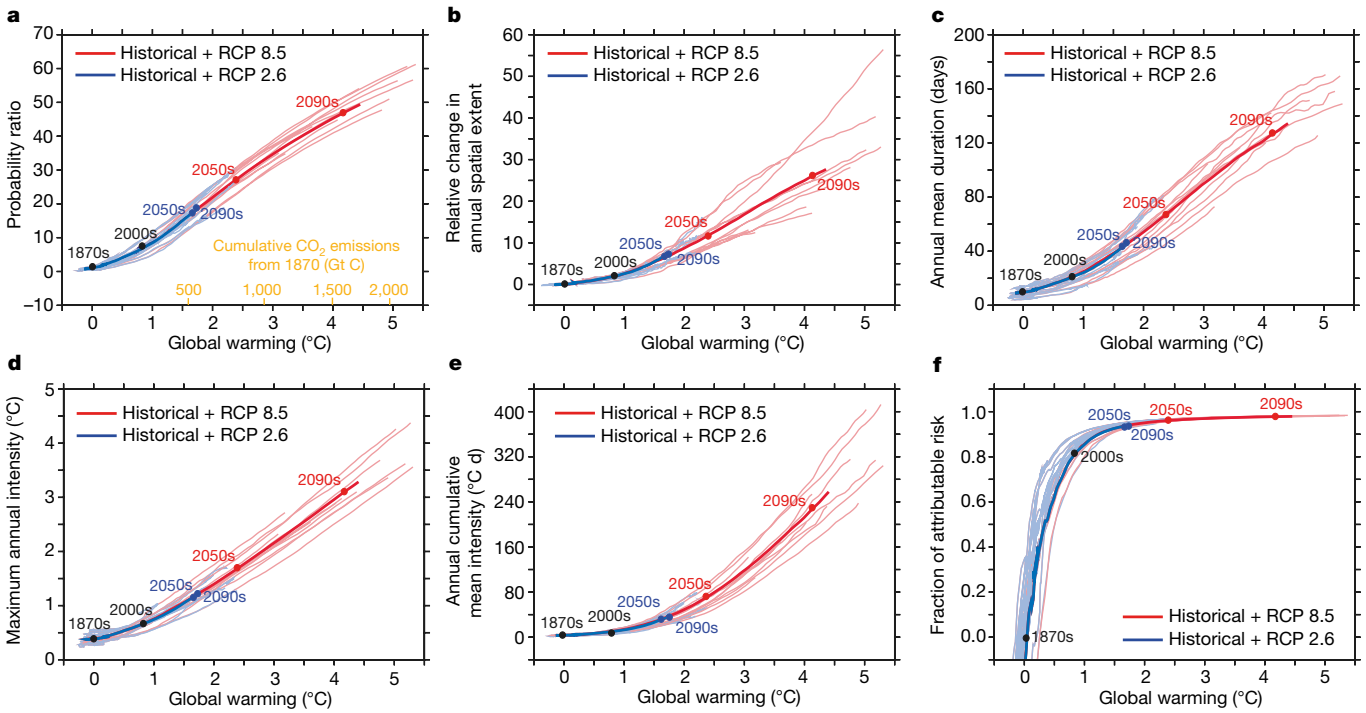


Fig. 1 | Simulated changes in MHW characteristics for different levels of global warming. a, **c**–**f**, Results are shown for the global aggregated annual mean probability ratio (**a**), duration (**c**), maximum intensity (**d**), cumulative mean intensity (**e**) and fraction of attributable risk (**f**) of MHWs exceeding the 99th preindustrial percentile. **b**, Ratio of the mean spatial extent at global warming conditions to that at 1861–1880 conditions. In all panels, the simulated MHW characteristics are plotted against simulated global mean atmospheric surface temperature changes since 1861–1880. The thinner lines represent individual model projections, whereas the thicker lines represent multi-model averages for the RCP 8.5 and RCP 2.6 scenarios. For all models, the historical simulations are merged with the RCP 2.6 and RCP 8.5 simulations. The time series are smoothed with a 20-year running mean and the year labels represent the central year of two decades. The cumulative CO_2 emissions (orange; in gigatons of C) corresponding to different global warming levels are shown in **a**, approximated using the RCP 8.5 ensemble average (see Methods).

[1]Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland. [2]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland. [3]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland. [4]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland. *e-mail: froelicher@climate.unibe.ch

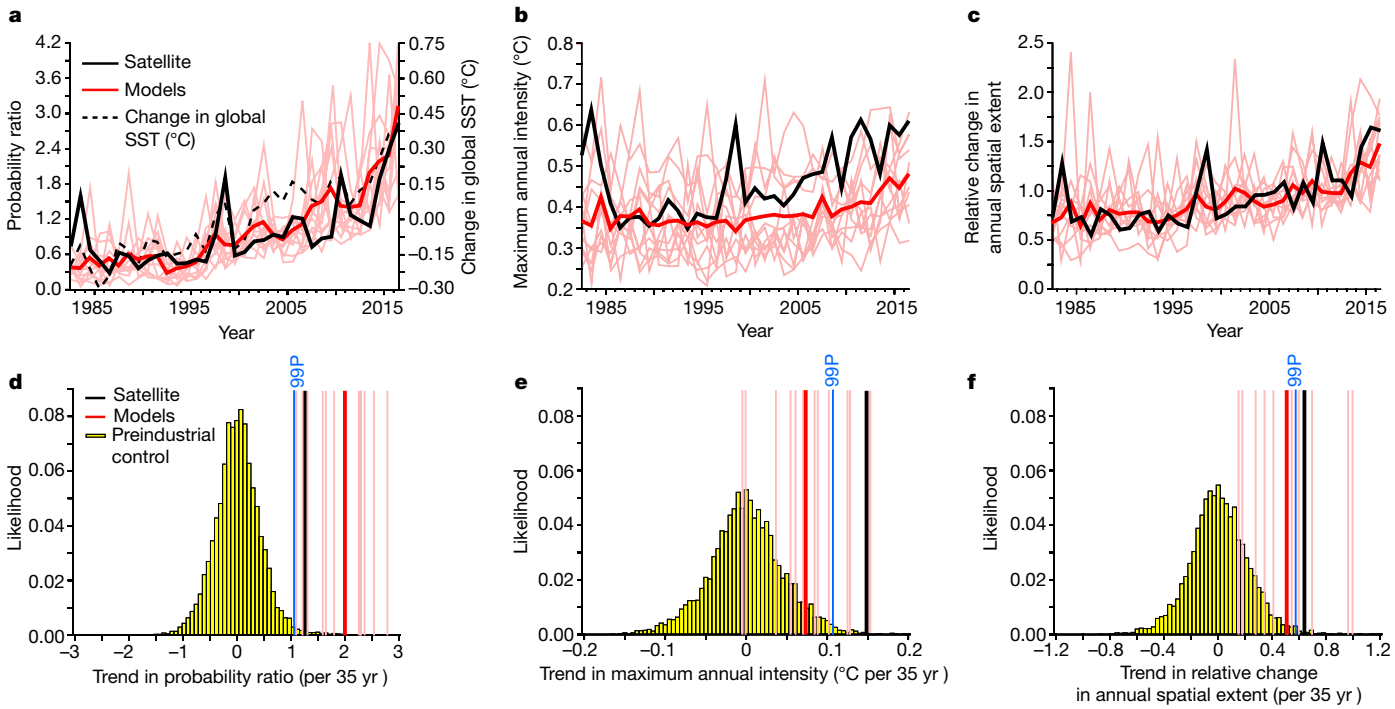


Fig. 2 | Observed and modelled trends in MHW characteristics over the satellite data taking period. a–c, The black lines show the observed changes in the global aggregated probability ratio (**a**), maximum annual intensity (**b**) and the ratio of the annual spatial extent at different years to that at 1982–2016 conditions (**c**) of MHWs exceeding the 1982–2016 99th percentile. The thick red lines indicate the simulated multi-model mean changes and the thin red lines the individual models of MHWs exceeding the 1982–2016 99th percentile. The observed global mean SST changes since 1982–2016 are shown in **a** as a black dashed line. **d–f**, The histograms show simulated 35-year trends of MHW characteristics in the preindustrial control simulations (see Methods for calculation details). The black and red vertical lines show the 35-year observed and simulated trends in 1982–2016 of MHWs exceeding the 1982–2016 99th percentile, and the blue vertical lines show the 99th percentile (labelled as '99P') of the probability density distribution of the preindustrial control simulation trends. The relative changes in the annual spatial extent are calculated as the ratio between the actual mean spatial extent and the average over the 1982–2016 period. Only simulations following the RCP 8.5 scenario are considered here because they best represent observed greenhouse gas emissions since 2006.

the associated impacts are. Although there is a rapidly growing literature on individual events[3,5,8,10,12], the underlying drivers and the degree to which they can be attributed to global warming[10,18] are currently not well known. This knowledge gap is of considerable concern given the high vulnerability of marine ecosystems and fisheries, but also human societies, to such events[19].

One of the first documented impacts of an MHW was the Mediterranean Sea heatwave event in 2003, which led to extensive mortality of benthic marine communities[11]. Other prominent examples are the record-high ocean warming off the coast of Western Australia in early 2011[20], the 2012 MHW in the northwest Atlantic[12], the persistent 2013–2015 extreme warm anomaly of the northeastern Pacific[21] and the 2015/2016 record-warm anomaly across most of the tropical and extratropical oceans[22]. MHWs have caused changes in biological production, toxic algal blooms[7], regime shifts in reef communities[4,8], mass coral bleaching[9] and mortalities of commercially important fish species[13], with cascading impacts on economies and societies[12].

Here, we detect past changes and assess future ones in different MHW characteristics using (i) remotely sensed daily global sea surface temperature (SST) data[22] covering the period 1982–2016, and (ii) daily output from twelve fully coupled global Earth system models (ESMs) covering the period 1861–2100 (see Methods). We identify an event as an MHW when the SST exceeds its local 99th percentile, as determined from daily data from either preindustrial model output or from satellite-based observations and model output over the 1982–2016 period. We then quantify the annual mean probability ratio (the fraction by which the number of MHW days per year has changed), relative change in the annual spatial extent (the average area of an individual heatwave), maximum annual intensity (maximum exceedance of the 99th percentile), annual mean duration (number of days of exceedance) and annual cumulative mean intensity (the product of the duration and the mean intensity of exceedance). We analyse three distinct periods: the preindustrial period (Fig. 1), the satellite data taking period (1982–2016; Fig. 2) and the future (Figs. 1, 3). We focus on summertime MHWs (that is, hottest days of the year), as many biological processes depend on the absolute temperature. The definition of MHWs needs to be altered, however, when MHWs in colder months can have an impact on biological processes[23].

In preindustrial times, the ESMs suggest that a typical MHW (with reference to preindustrial climatology) lasted 11 days (intermodel range, 6–14 days), had an intensity of up to 0.4 °C (0.3–0.5 °C) and a cumulative mean intensity of 3 °C d (2–4 °C d) (Fig. 1, Extended Data Table 1). MHWs occur coherently with a typical spatial extent of 4.2×10^5 km^2 (1.2×10^5–7.0×10^5 km^2). Under the present-day 1 °C global warming scenario, these models project a nine-fold (6–12) increase in the probability of occurrence of an MHW and a three-fold (1–3) increase in its spatial extent. Further, they project that the duration and the maximum annual and cumulative mean intensity have increased to 25 days (15–33 days), 0.8 °C (0.6–1 °C) and 13 °C d (8–18 °C d), respectively.

These century-scale changes can be put into perspective by determining the trend that they imply over the 35-year period 1982–2016, for which we have satellite observations. To this end, we change the reference for the definition of MHWs to this period. This has virtually no impact on trend computation, but affects the magnitude of the MHW characteristics. Over these 35 years, the models simulate mean changes in the probability ratio and maximum annual intensity of +2.0 (1.1–2.8), +0.07 °C (−0.01 °C to 0.15 °C), respectively, and relative changes in the annual spatial extent of +0.53 (0.17–1.00) (thick red lines in Fig. 2). These multi-model mean trends are at the high end or outside the range of those expected from internal variability (histograms in Fig. 2d–f), which is determined from the preindustrial

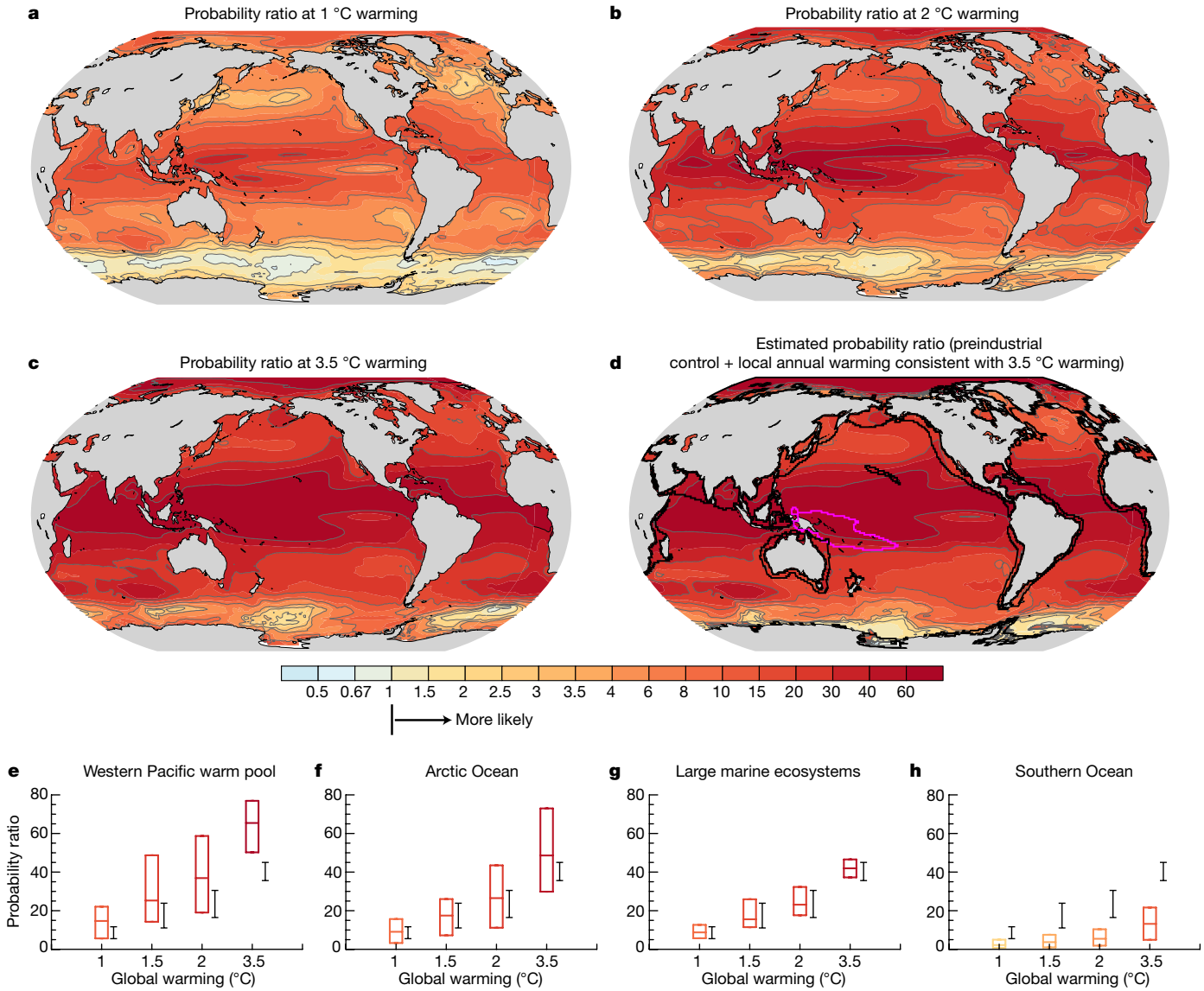


Fig. 3 | Regional changes in probability of MHW days for different global warming levels. a–d, Changes in the probability of MHW days exceeding the preindustrial 99th percentile for a global warming level of 1 °C (**a**), 2 °C (**b**) and 3.5 °C (**c**). To show that the occurrence of MHWs is mainly driven by a simple shift of the whole temperature distribution, in **d** we have added the local annual SST change that is consistent with a 3.5 °C global warming to the preindustrial SST distribution. **e–h**, Changes are regionally aggregated over the western Pacific warm pool (**e**), the Arctic Ocean at >75° N (**f**), large marine ecosystems (**g**) and the Southern Ocean at 45° S–65° S (**h**). Box plots indicate the multi-model mean, minimum and maximum changes in probability, and their colour indicates the value of the probability ratio, using the same colour coding as for **a–d**. Black bars represent the multi-model minimum and maximum of the global averaged probability changes. In **d**, the western Pacific warm pool region is highlighted by a purple solid contour and the large marine ecosystems are shown with black contours adjacent to the continents in coastal waters. The grey contours in **a–d** highlight pattern structures. The large marine ecosystems provide 95% of the world's annual marine fishery yields[30] and have been developed to enable ecosystem-based marine resource management within ecologically bounded transnational areas. The maps were created using the NCAR Command Language (https://www.ncl.ucar.edu).

control simulations. This indicates that the climate change signal could be strong enough to be detected in observations.

The corresponding 35-year trends in the satellite observations (thick black lines in Fig. 2d–f) are of similar magnitudes as the simulated ones (red lines in Fig. 2d–f). The observations reveal a significant increase in the probability ratio ($+1.29 \pm 0.28$ per 35 years; $P < 0.01$ using a two-sided t-test), maximum intensity ($+0.15 \pm 0.05$ °C per 35 years; $P < 0.01$) and spatial extent ($+0.66 \pm 0.13$ per 35 years; $P < 0.01$) (thick black lines in Fig. 2). These observed trends are statistically significantly outside the model-based estimate of trend variability arising from internal variability, but are within the simulated intermodel uncertainty for trends arising from simulations that include anthropogenic forcing (thin red lines in Fig. 2d–f). Assuming that the model-based estimate of internal variability is accurate, we can conclude with high confidence that the observed trends in the MHW days, maximum intensity and spatial extent of MHWs are largely caused by long-term ocean warming. Support for this conclusion comes from the fact that SST variations have also a large effect on the year-to-year variability of the different MHW characteristics. The observed temporal evolution of the annual mean SST (black dashed line in Fig. 2a) has strong correlations with the probability ratio ($r^2 = 0.66$ for global SST and probability ratio) and the spatial extent ($r^2 = 0.65$) of MHWs, but relatively weak correlation with their maximal intensity ($r^2 = 0.36$).

The satellite records allow us also to assess the characteristics of the modelled MHWs, allowing us to establish confidence levels for the projections. The modelled spatial pattern in the probability ratio, maximum intensity and the frequency distribution of the spatial extent of MHWs are comparable to the observed ones over the satellite data

taking period (Extended Data Figs. 1, 2), giving us confidence in the corresponding projections. By contrast, the duration, cumulative mean intensity and absolute spatial extent of the MHWs are less well captured by the models, with substantial biases in the corresponding patterns (Extended Data Figs. 1, 2). This indicates that we need to be more careful when interpreting the modelled changes of these characteristics.

For the future and all ocean basins, the ESMs project more frequent, extensive, intense and longer-lasting MHWs (Fig. 1a–e, Fig. 3, Extended Data Tables 1, 2 and Extended Data Fig. 3) (here the reference period is set back to preindustrial times). The magnitude of these changes scales with the global mean temperature and the cumulative CO_2 emissions that drive this global warming (Fig. 1). This scaling is independent of the warming path, that is, it does not depend on whether a particular warming is reached sooner (RCP 8.5, high-emission scenario; see Methods) or later (RCP 2.6, low-emission scenario compatible with the Paris Agreement). It also does not depend on the reference period, as the use of the satellite reference period would only shift this curve slightly to the left (Extended Data Fig. 4a). This allows us to assess the future projections in terms of warming levels rather than the time when this warming is reached.

For 3.5 °C warming, the probability of occurrence of an MHW is 41 times (intermodel range, 36–45 times) higher than in preindustrial times (Fig. 1a, Extended Data Table 2). In other words, a one-in-a-hundred-days event at preindustrial levels is projected to become a one-in-three-days event at this level of global warming. The spatial extent of the annual mean is projected to become 21 (15–29) times larger, its duration to increase to 112 days (92–129 days), and its maximum intensity to rise to 2.5 °C (2.1–2.9 °C) (Fig. 1b–d, Extended Data Table 1). The projected increase in maximum intensity is smaller than the increase in global mean temperature owing to the substantially lower rate of warming by the surface ocean compared to land. The increase in the duration and intensity also leads to a strong increase in the cumulative mean intensity of MHWs of 164 °C d (126–214 °C d) (Fig. 1e, Extended Data Table 1).

These large increases in the different MHW characteristics are substantially reduced if warming is kept below 2 °C, or even below 1.5 °C. The probability of occurrence for an MHW under the 1.5 °C warming scenario is only 40% of that under 3.5 °C warming. The relative change in the spatial extent of a typical MHW would be 25%, the duration 35% and the maximum intensity 45% of those at 3.5 °C.

The probability of MHWs is projected to increase almost everywhere, and the increase is largest in the tropics and the Arctic Ocean and smallest in the Southern Ocean (Fig. 3). The main reason for the large changes in probability in the tropics, and especially in the western Pacific warm pool, is the small variations in SST in these areas, both seasonally and from year to year[24]. As a result, the same changes in annual mean SST lead to much larger changes in the probability of exceeding the 99th percentile. The same applies to the Arctic Ocean, where SST variations below year-round sea ice are very small[25]. This is in contrast to the Southern Ocean, where surface waters are projected to stay relatively cool, and therefore the probability ratio does not increase much under all warming levels. The projected increase in the probability of MHWs in the coastal large marine ecosystems (indicated as black coastal regions in Fig. 3d) has similar magnitude to the global increase under 2 °C warming.

Because of the large increase in the probability ratio with warming, the simulated fraction of attributable risk—that is, the anthropogenic contribution to the probability of an event—reaches 0.87 (0.78–0.91) already under a present-day level of 1 °C warming (Fig. 1f, Extended Data Table 2). This implies (under the assumption that the models simulate naturally occurring MHWs with fidelity) that 87% of the currently occurring MHWs (defined relative to preindustrial conditions) can be attributed to global warming. Because this warming is primarily driven by anthropogenic emissions of greenhouse gases[26], there is a direct link between human action and the simulated increase in MHWs; this supports our conclusion drawn from the satellite data. Clearly, any specific MHW event still arises from the natural variability in the climate system, but the present-day level of global warming has substantially increased the odds of an MHW to occur. The simulated fraction of attributable risk approaches unity (0.94–0.97) already at 2 °C, implying that essentially all MHWs are due to anthropogenic warming at this or higher levels of warming.

The changes in the occurrence of MHWs are mainly driven by the global-scale shift in mean SSTs. We demonstrate this by adding the simulated spatial warming pattern that is consistent with a global warming of 3.5 °C to the results from the preindustrial control run (Fig. 3d). This yields probability ratio values and patterns that are similar to the results from the transient simulations. It also implies that changing the reference period would not change the relationship between the different MHW characteristics and the amount of warming (Extended Data Fig. 4a). A notable exception is the northern Arctic Ocean, where the SST remains close to freezing temperature during boreal winter months even under the RCP 8.5 scenario[25]. This slightly damps the increase in the probability ratio that would be expected from a global-scale shift in the mean SST.

An important assumption in our analyses is that the employed ESMs simulate MHWs in a sufficiently realistic manner. We consider our results for the probability ratio, maximum intensity and the relative changes in the spatial extent of MHWs to be robust, especially given the good agreement with observations (Fig. 2 and Extended Data Fig. 1) and the relatively small intermodel spread in MHW projections. However, the simulated MHWs last generally longer and are spatially more extensive than observed ones (Extended Data Fig. 1, 2), which is probably caused by the relatively coarse resolution of the ESMs. High-resolution coupled models are needed to resolve mesoscale processes in the atmosphere and the ocean that may be critical to improve the representation of the duration and spatial extent of MHWs. In addition, the conclusion that global warming will lead to a strong increase in all MHW characteristics does not depend on how an MHW is defined (Extended Data Fig. 5), but the quantitative results of MHWs can vary substantially with that definition.

An increase in MHWs will probably increase the risk of severe, pervasive and long-lasting impacts on marine organisms[8], especially on those with reduced mobility and high vulnerability, such as coral reefs, and those living at low latitudes, where many marine species live close to their upper thermal limits[27]. However, the responses of marine organisms and ecosystems to MHWs can be variable and difficult to predict owing to species- and system-specific responses[28,29]. Therefore, better understanding of the response of marine organisms and ecosystems to MHWs and extreme events in other stressors is urgently needed to assess the full risk for marine organisms and ecosystems under global warming.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0383-9.

Received: 29 November 2017; Accepted: 15 May 2018;
Published online 15 August 2018.

1. Hobday, A. J. et al. A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr.* **141**, 227–238 (2016).
2. Scannell, H. A., Pershing, A. J., Alexander, M. A., Thomas, A. C. & Mills, K. E. Frequency of marine heatwaves in the North Atlantic and North Pacific since 1950. *Geophys. Res. Lett.* **43**, 2069–2076 (2016).
3. Marbà, N. & Duarte, C. M. Mediterranean warming triggers seagrass (Posidonia oceanica) shoot mortality. *Glob. Change Biol.* **16**, 2366–2375 (2010).
4. Wernberg, T. et al. An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nat. Clim. Change* **3**, 78–82 (2013).
5. Di Lorenzo, E. & Mantua, N. Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nat. Clim. Change* **6**, 1042–1047 (2016).
6. McCabe, R. M. et al. An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions. *Geophys. Res. Lett.* **43**, 10,366–10,376 (2016).
7. Cavole, L. et al. Biological impacts of the 2013–2015 warm-water anomaly in the Northeast Pacific: winners, losers, and the future. *Oceanography* **29**, 273–285 (2016).

8. Wernberg, T. et al. Climate-driven regime shift of a temperate marine ecosystem. *Science* **353**, 169–172 (2016).
9. Hughes, T. P. et al. Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
10. Oliver, E. C. J. et al. The unprecedented 2015/16 Tasman Sea marine heatwave. *Nat. Commun.* **8**, 16101 (2017).
11. Garrabou, J. et al. Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave. *Glob. Change Biol.* **15**, 1090–1103 (2009).
12. Mills, K. E. et al. Fisheries management in a changing climate: lessons form the 2012 ocean heat wave in the Northwest Atlantic. *Oceanography* **26**, 191–195 (2014).
13. Caputi, N. et al. Management adaptation of invertebrate fisheries to an extreme marine heat wave event at a global warming hot spot. *Ecol. Evol.* **6**, 3583–3593 (2016).
14. Pershing, A. J. et al. Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science* **350**, 809–812 (2015).
15. Oliver, E. C. J. et al. Longer and more frequent marine heatwaves over the past century. *Nat. Commun.* **9**, 1324 (2018).
16. Climate Action Tracker. *Improvement in warming outlook as India and China move ahead, but Paris Agreement gap still looms large* (2017); https://climateactiontracker.org/documents/61/CAT_2017-11-15_ImprovementInWarmingOutlook_BriefingPaper.pdf.
17. Pachauri, R. K. & Meyer, L. A. (eds.) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, Geneva, 2014).
18. Walsh, J. E. et al. The high latitude marine heat wave of 2016 and its impacts on Alaska. *Bull. Am. Meteorol. Soc.* **99**, 39–43 (2018).
19. Frölicher, T. L. & Laufkötter, C. Emerging risks from marine heat waves. *Nat. Commun.* **9**, 650 (2018).
20. Pearce, A. F. & Feng, M. The rise and fall of the 'marine heat wave' off Western Australia during the summer of 2010/2011. *J. Mar. Syst.* **111–112**, 139–156 (2013).
21. Bond, N. A., Cronin, M. F., Freeland, H. & Mantua, N. Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophys. Res. Lett.* **42**, 3414–3420 (2015).
22. Banzon, V., Smith, T. M., Liu, C. & Hankins, W. A long-term record of blended satellite and in situ sea surface temperature for climate monitoring, modeling and environmental studies. *Earth Syst. Sci. Data* **8**, 165–176 (2016).
23. Morley, J. W., Batt, R. D. & Pinsky, M. L. Marine assemblages respond rapidly to winter climate variability. *Glob. Change Biol.* **23**, 2590–2601 (2017).
24. Frölicher, T. L., Rodgers, K. B., Stock, C. A. & Cheung, W. W. L. Sources of uncertainties in 21st century projections of potential ocean ecosystem stressors. *Glob. Biogeochem. Cycles* **30**, 1224–1243 (2016).
25. Carton, J. A., Ding, Y. & Arrigo, K. R. The seasonal cycle of the Arctic Ocean under climate change. *Geophys. Res. Lett.* **42**, 7681–7686 (2015).
26. Bindoff, N. L. et al. *Detection and Attribution of Climate Change: from Global to Regional. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, 2013).
27. Comte, L. & Olden, J. D. Climatic vulnerability of the world's freshwater and marine fishes. *Nat. Clim. Change* **7**, 718–722 (2017).
28. Logan, C. A., Dunne, J. P., Eakin, C. M. & Donner, S. D. Incorporating adaptive responses into future projections of coral bleaching. *Glob. Change Biol.* **20**, 125–139 (2014).
29. Donelson, J. M., Munday, P. L., McCormick, M. I. & Pitcher, C. R. Rapid transgenerational acclimation of a tropical reef fish to climate change. *Nat. Clim. Change* **2**, 30–32 (2012).
30. Stock, C. A. et al. Reconciling fisheries catch and ocean productivity. *Proc. Natl Acad. Sci. USA* **114**, E1441–E1449 (2017).

Acknowledgements T.L.F. acknowledges support from the Swiss National Science Foundation under grant PP00P2_170687 and N.G. under grant XEBUS 200020_175787. E.M.F and N.G. acknowledge support from ETH Zürich. This work is part of the Nippon Foundation Nereus Program, a collaborative initiative by the Nippon Foundation and partners, including ETH Zürich. We thank the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP5, and the climate modelling groups (listed in Methods) for producing and making available their model output. We thank T. Stocker, F. Joos and C. Raible for discussions, U. Beyerle for downloading most of the CMIP5 data, and D. Kessler and M. Aschwanden for their initial analyses. The maps in Fig. 3 and Extended Data Figs. 1– 3 were created using the NCAR Command Language (https://www.ncl.ucar.edu).

Reviewer information *Nature* thanks R. Asch, E. Oliver and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions All authors designed the study. T.L.F. performed the analysis and wrote the manuscript together with N.G., with substantial input from E.M.F.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0383-9.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.L.F.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

We analyse daily SST and surface atmospheric temperature data from simulations (using the first ensemble member, r1i1p1) of twelve coupled ESMs that were considered in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) and for which the output necessary to analyse changes in daily SST was available (Extended Data Table 3). All model simulations were run over the historical 1861–2005 period and over the 2006–2100 period, following both a high-emission scenario (RCP 8.5; RCP, representative concentration pathway) and a low-emission scenario compatible with the Paris Agreement (RCP 2.6).

In addition, we use the National Oceanic and Atmospheric Administration's ¼° daily optimum interpolation SST dataset[22,31] version 2.0, obtained by the Advanced Very High Resolution Radiometer and covering the period 1 January 1982 to 31 December 2016 (www.ncdc.noaa.gov/oisst/; accessed on 6 July 2017). The dataset combines observations from different platforms, such as satellites, ships and buoys, and includes bias adjustment of satellite and ship observations to compensate for platform differences and sensor biases. For comparison with the coarse-resolution models, the 0.25° × 0.25° satellite data were regridded daily onto a regular 1° × 1° grid by averaging over the 1°-grid cells before calculating the characteristics of the MHW.

We define an event as an MHW when the daily SST exceeds the 99th percentile (a one-in-a-hundred-days event). We test the sensitivity of the results by also using the 90th (a one-in-ten-days event), the 99.9th (a one-in-2.74-years event) and the 99.99th (a one-in-27.4-years event) percentiles (Extended Data Fig. 5). The percentiles are calculated for each grid cell from multi-centennial preindustrial control simulations (most simulations are for 500 years or longer). This ensures that even the local 99.99th percentile is well defined. The same preindustrial control simulation is used to define the reference global mean temperature relative to which the warming targets are computed. Changing the reference period to present-day (that is, 2007–2026; ±10 years centred on today) would just shift the values on the *x* axis in all panels of Fig. 1, but would not change the relationship between the different MHW characteristics and the amount of global warming (Extended Data Fig. 4a). Because some models have constant year-round SSTs in a few grid cells under sea ice in the preindustrial control simulations, grid cells in which the average yearly number of MHWs over the entire control simulation deviates by more than 5% from the theoretical number (for example, 3.65 days for the 99th percentile) are masked out. For the analysis of atmospheric heatwaves over land, we use the same definition as for MHWs. For analysis over the satellite data taking period, we use the entire 1982–2016 period as the baseline period for both the models and the satellite data.

The usage of a percentile-based threshold allows the quantification of MHWs across locations that differ in variability. An absolute threshold would only be relevant in terms of impacts in some regions but not in others. By using percentile-based characteristics, no assumption is made regarding the underlying probability temperature distribution, and potential model-observation biases in the mean and higher-order statistical moments of the probability temperature distribution are implicitly taken into account. This increases our confidence in the simulated probability ratio, but the simulated spatial extent and duration (and intensity) of MHWs may still differ from observations. Our definition differs from that proposed by Hobday et al.[2], who define an MHW by using a much lower seasonally varying percentile threshold (90th rather than 99th), but impose a duration of at least five days. Relative to the results obtained with our definition, the definition of Hobday et al. would lead to an increase of the number of heatwave days, including the cold seasons, because the vast majority of our heatwaves last longer than five days. However, using their definition would not change our conclusion about the robust increase in all MHW metrics under global warming, because this result is essentially insensitive to the percentile threshold that we choose (Extended Data Fig. 5).

For each MHW, we calculate a series of characteristics, such as the duration (in days; number of days of percentile threshold exceedance), the maximum intensity (in °C; maximum SST anomaly with respect to the percentile threshold over the duration of the heatwave), the spatial extent (in km^4), the cumulative mean intensity (in °C d; the mean intensity multiplied with the duration of an event), and the probability ratio, $PR = P_1/P_0$, where P_1 is the probability of exceeding a relative threshold at any given point in time (for example, today) and P_0 the probability of exceeding that threshold during the preindustrial control or satellite climatological period. The cumulative mean intensity may indicate the integrated impact of an MHW on an organism's health—a similar measure, the degree heating days or weeks, is commonly used to identify areas where substantial coral bleaching is likely to occur[32]. We then calculate annual statistics, including the number of MHW days per year, the changes in the annual averaged spatial extent of an MHW relative to 1861–1880 or the satellite climatological period, the annual mean duration of single contiguous MHW events in a given year, the maximum annual intensity, the annual mean cumulative mean intensity and the annual mean fraction of attributable risk ($FAR = 1 - P_0/P_1$)[10,18,33–36].

Because the observed global warming primarily results from human influence, we can attribute the changes in the occurrence of MHWs to human-induced global warming[33]. The FAR was initially introduced to represent a fraction of the probability of individual observed events[33,37]. Here, we extend the FAR framework to the global scale[34,38,39] to represent the probability for a class of events exceeding a certain threshold over the globe. For a given MHW, the probability ratio can be interpreted as a change in the odds of the occurrence of local SST anomalies exceeding a certain local threshold. The regional or global aggregated probability ratio expresses a change in the global occurrence of SST anomalies exceeding local thresholds.

We calculate the MHW properties by using slightly different frameworks. The probability ratio, maximum intensity, duration, cumulative mean intensity and FAR are defined when local (grid-cell) SST exceeds the local 99th percentile and where adjacent grid cells can have different values. The intensity and duration refers to the properties of a contiguous event, but the probability ratio refers to MHW days per year, regardless of how they are distributed across different events. For the spatial extent, we aggregate adjacent grid cells that are above the 99th percentile together to form a single event. To calculate the spatial extent of individual MHWs, we isolate the individual MHWs per day using the function *skimage.measure.label* of the Python image processing tool scikit-image. The global estimate of these characteristics is calculated with an area-weighted average across all ocean grid points in each year from 1861 to 2100. All MHW characteristics are calculated on the native model grid, which differs in resolution across the models, but multi-model means and globally aggregated characteristics are calculated and shown on a regular 1° × 1° grid.

We usually express the changes in MHW characteristics as changes for particular global warming levels (that is, 1 °C, 2 °C and 3.5 °C). These global warming levels are calculated for each model and scenario individually by subtracting the simulated global annual mean atmospheric surface temperature, averaged over the 20-year period centred around the year when the respective global warming level is reached, from the simulated global mean atmospheric surface temperature averaged over the 1861–1880 period.

The cumulative CO_2 emissions corresponding to the different global warming levels (orange ticks on horizontal axis in Fig. 1a) are approximated using calculated cumulative CO_2 emissions from the RCP 8.5 average of eight models for which necessary data were available (Extended Data Table 3). This means that 500 Gt C corresponds to 1.6 °C, 1,000 Gt C to 2.8 °C, 1,500 Gt C to 4.0 °C and 2,000 Gt C to 4.9 °C. No uncertainties are assigned to these values. We note that these eight models as a class have a relatively low transient climate response to cumulative carbon emissions, and therefore cumulative carbon emission estimates for a certain global warming level are relatively large[40].

To test whether the observed multi-decadal trends over the satellite data taking period are different from what would be expected from internal variability, we compare the observed global aggregated trends in the probability ratio, maximum intensity and spatial extent with the probability density function of 35-year-long trends derived from the 12 multi-century control simulations of the different ESMs. In total, we calculated 6,460 35-year trends.

We also used a ten-member ensemble simulation of the NCAR-DOE CESM model[34] to show that internal variability may induce uncertainty at the local level, but plays a negligible role in explaining the global changes in the different MHW characteristics, as all ten realizations yield very similar results (Extended Data Fig. 4). We also show that our simulated changes in the MHW characteristics do not depend on the choice of calculation method for the climatological 99th percentiles. In fact, the simulated changes in the MHW characteristics are similar when determining the local 99th percentiles from a simulation of the GFDL ESM2M model forced with observed solar and volcanic boundary conditions but with greenhouse gases and aerosols concentration set to preindustrial levels (Extended Data Fig. 6). Only the NCAR-DOE CESM and GFDL ESM2M models provide the daily output necessary to analyse the sensitivity of the results to internal variability and to the calculation of the climatological baseline period.

We used the western equatorial Pacific biome definition of Fay and McKinley[41] to highlight the western Pacific warm pool region in Fig. 3. The biogeographical biomes in ref. [41] are defined by distinct SSTs, maximum mixed-layer depths and summer chlorophyll concentrations, and capture patterns of large-scale biogeochemical function at the basin scales.

Under any level of warming, MHWs are projected to occur much more frequently than land-based heatwaves (Extended Data Table 2). The probability of MHW days is about 1.6 times higher than that for land-based heatwave days under 2 °C global warming, even though the global SSTs are projected to increase by only 0.55 °C per degree of surface air temperature warming over land (Extended Data Fig. 7). The larger increase in the probability ratio is obtained because the temperature variability over land is much larger than over the ocean[19], leading to a smaller signal-to-noise ratio. This is evidenced by the 8.0 °C difference between the 99th percentile and the annual mean air temperature averaged over the global

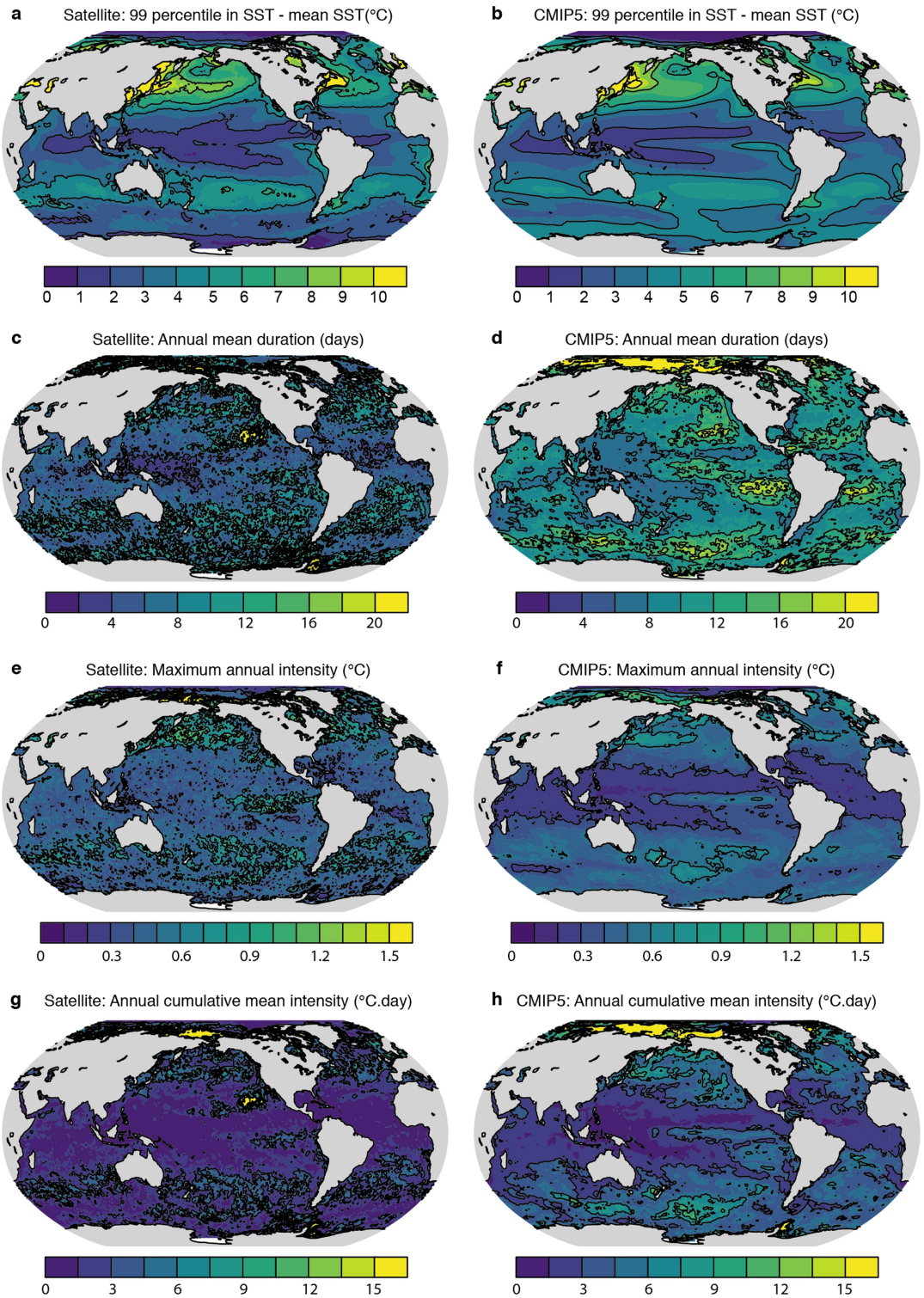
land surface at preindustrial times, which is much larger than the 3.7 °C difference over the ocean.

In general, the probability ratio, and therefore also the FAR, increase the most for very rare extremes (Extended Data Table 2); that is, they increase much more if MHWs are defined with more extreme preindustrial percentile thresholds (Extended Data Fig. 5). For example, the probability ratio is 9 (intermodal range, 6–12) for moderate MHWs (defined with the 99th preindustrial percentile) and 141 (47–296) for the rarest MHWs (99.99th preindustrial percentile) under 1 °C global warming.

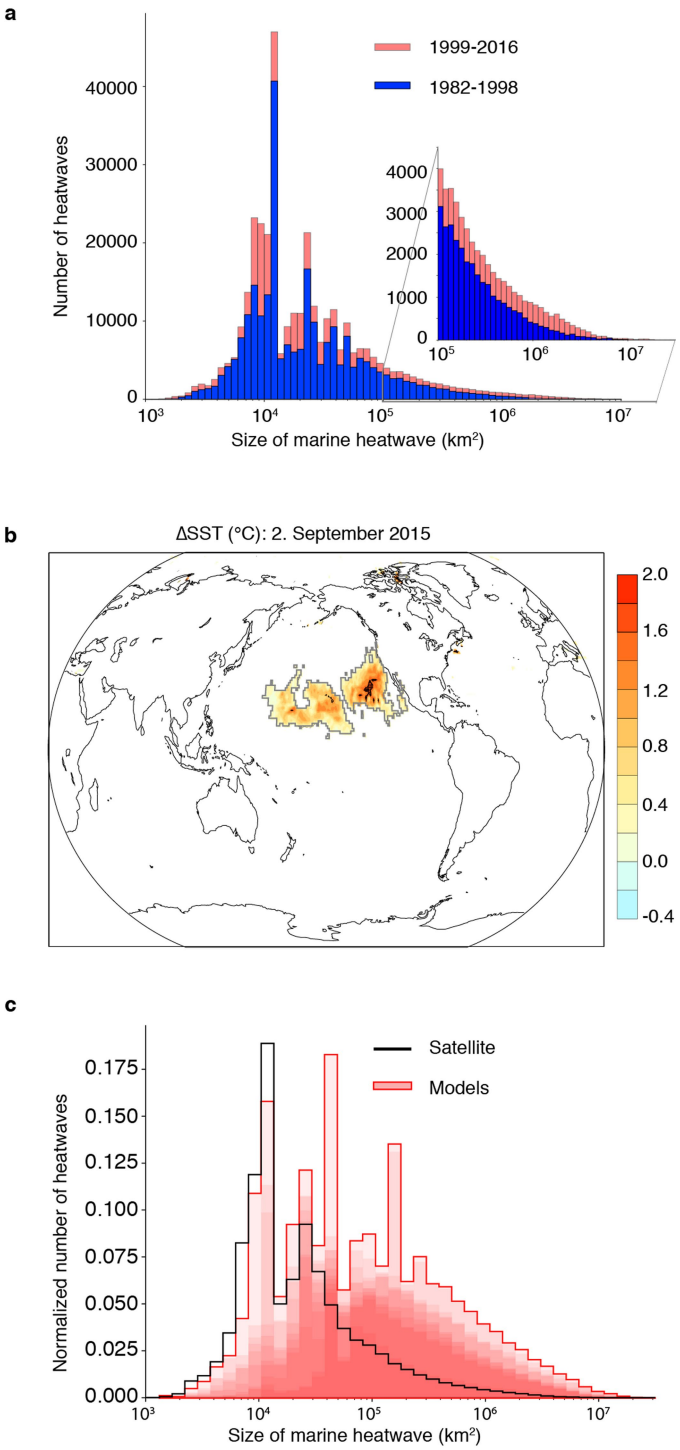
Code availability. The code used to produce the figures in this paper is available from the corresponding author upon request.

Data availability. The CMIP5 data used for this study can be accessed at http://pcmdi9.llnl.gov/ and the satellite SST observations can be accessed at www.ncdc.noaa.gov/oisst/. Other datasets generated during the current study are available from the corresponding author upon reasonable request.

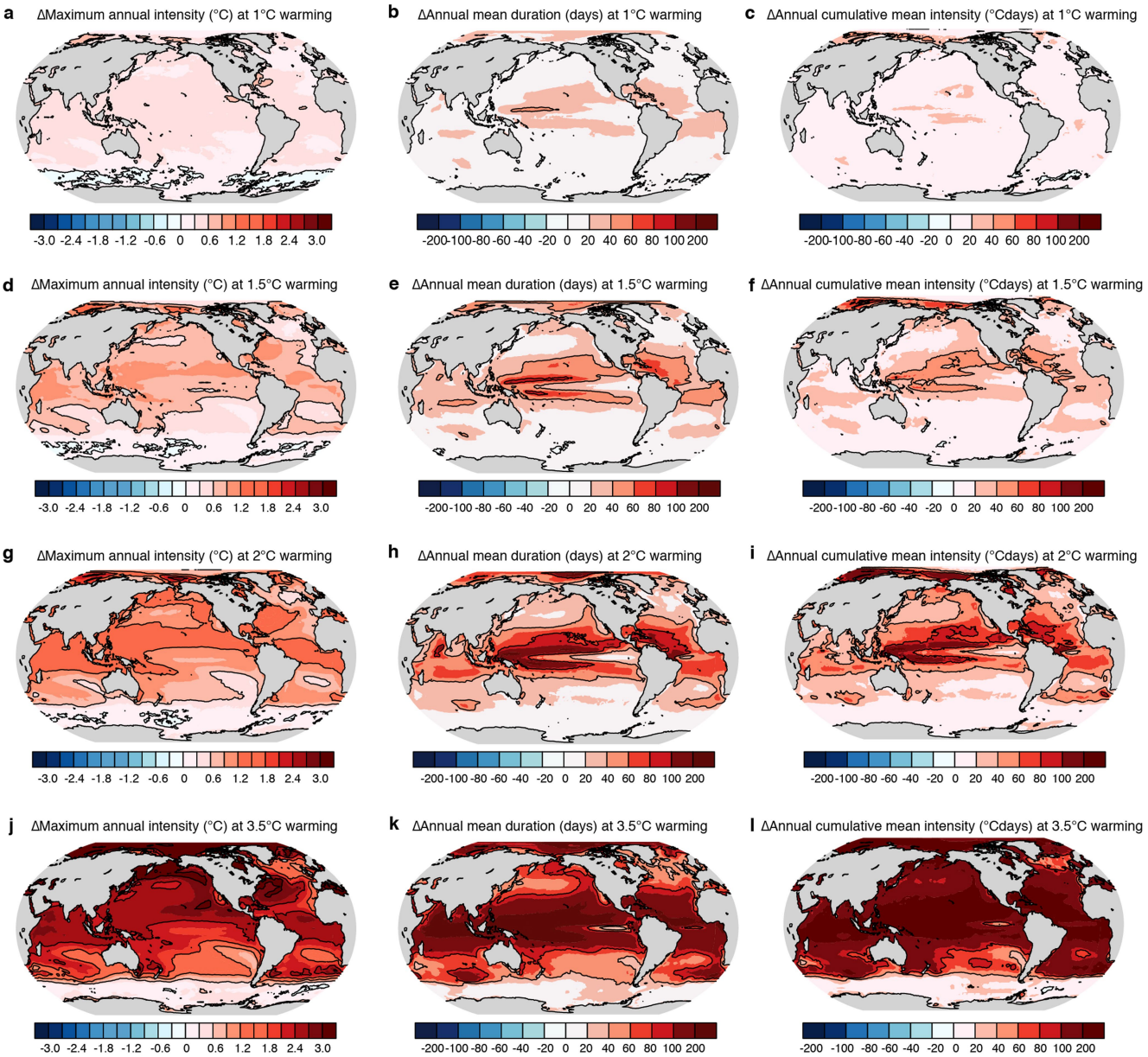
31. Reynolds, R. W. et al. Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* **20**, 5473–5496 (2007).
32. Eakin, C. M. et al. Caribbean corals in crisis: record thermal stress, bleaching, and mortality in 2005. *PLoS One* **5**, e13969 (2010).
33. Stott, P., Stone, D. & Allen, M. Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004); corrigendum **436**, 1200 (2005).
34. Fischer, E. M. & Knutti, R. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Clim. Change* **5**, 560–564 (2015).
35. Jacox, M. G., Alexander, M. A., Mantua, N. & Scott, J. D. Forcing of multiyear extreme ocean temperatures that impacted California Current marine resources in 2016. *Bull. Am. Meteorol. Soc.* **99**, 27–33 (2018).
36. Newman, M., Wittenberg, A. T., Cheng, L., Compo, G. P. & Smith, C. A. The extreme 2015/16 El Nino, in the context of historical climate variability and change. *Bull. Am. Meteorol. Soc.* **99**, 16–20 (2018).
37. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232 (2002); corrigendum **489**, 590 (2012).
38. Angélil, O. et al. Attribution of extreme weather to anthropogenic greenhouse gas emissions: Sensitivity to spatial and temporal scales. *Geophys. Res. Lett.* **41**, 2150–2155 (2014).
39. King, A. D. et al. Emergence of heat extremes attributable to anthropogenic influences. *Geophys. Res. Lett.* **43**, 3438–3443 (2016).
40. Frölicher, T. L. & Paynter, D. J. Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales. *Environ. Res. Lett.* **10**, 075002 (2015).
41. Fay, A. R. & McKinley, G. A. Global ocean biomes: mean and temporal variability. *Earth Syst. Sci. Data* **7**, 107–128 (2014).



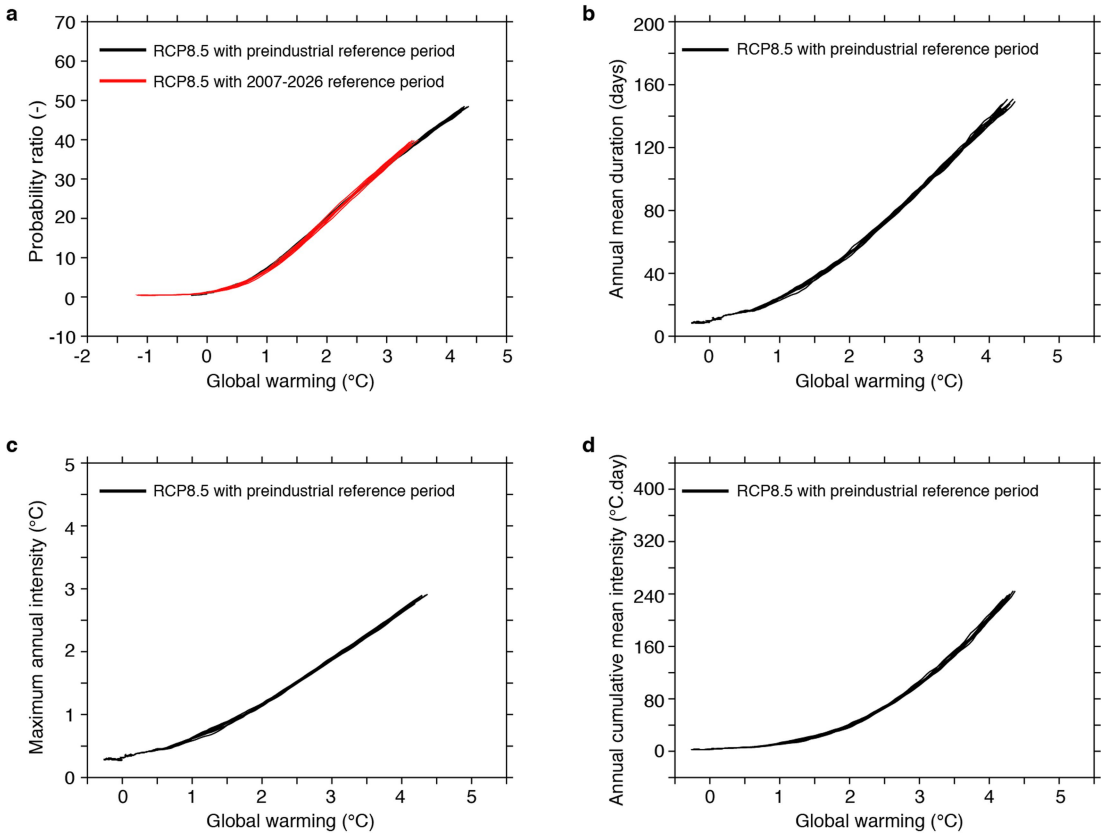
Extended Data Fig. 1 | Observed and simulated MHW characteristics exceeding the 1982–2016 99th percentile, averaged over the 1982–2016 period. **a**, **b**, Differences between the 99th percentile in SST and the annual mean SST. **c**, **d**, Annual mean duration of MHWs. **e**, **f**, Maximum annual intensity of MHWs. **g**, **h**, Annual cumulative mean intensity of MHWs. Satellite-derived patterns (**a**, **c**, **e**, **g**) and CMIP5 multi-model mean patterns (**b**, **d**, **f**, **h**). The black contours in all panels highlight the pattern structures. The spatial correlation between the CMIP5 multi-model mean and the satellite-based estimates is $r^2 = 0.80$ for **a** and **b**, $r^2 = 0.15$ for **c** and **d**, $r^2 = 0.43$ for **e** and **f**, and $r^2 = 0.18$ for **g** and **h**.



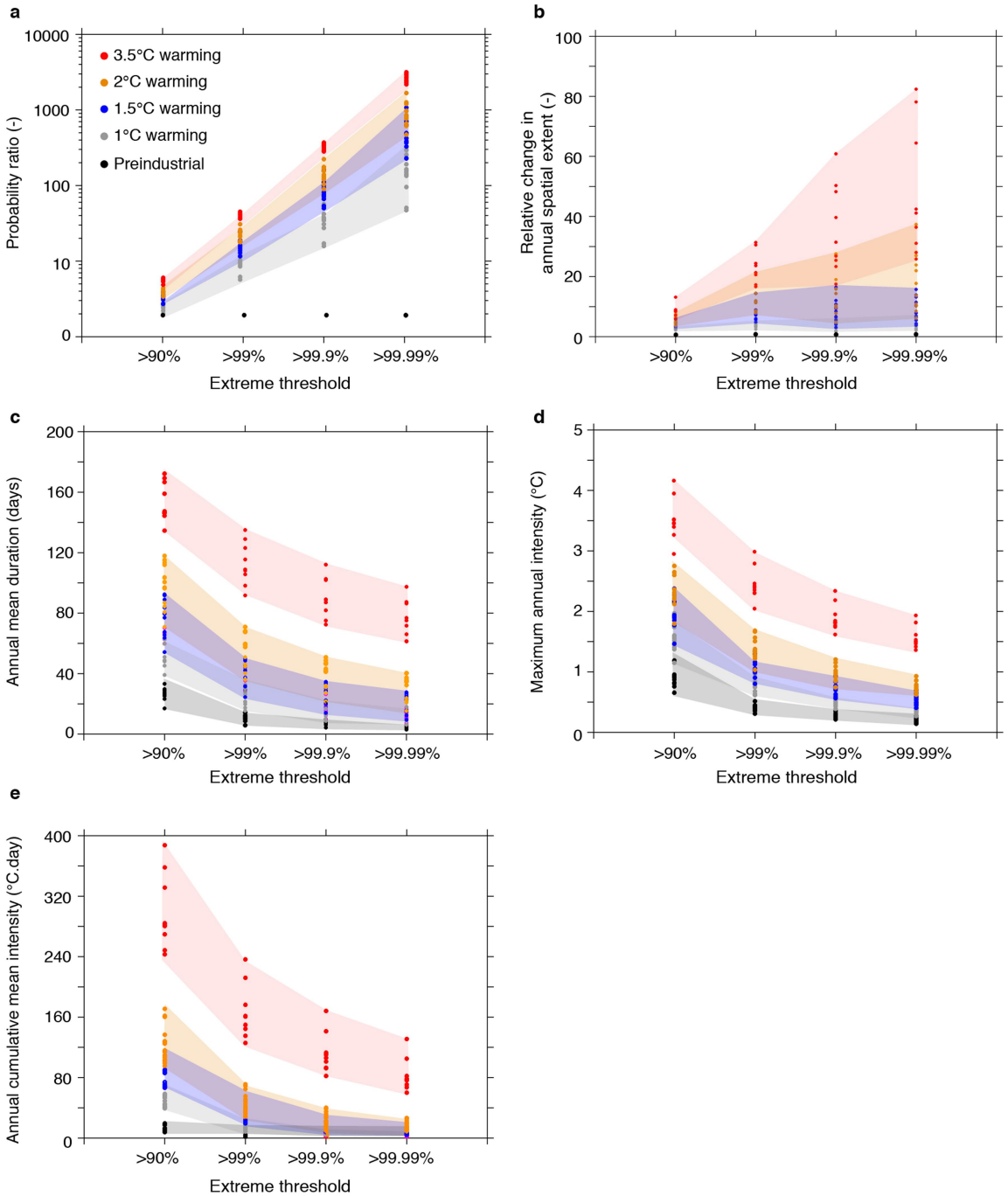
Extended Data Fig. 2 | Spatial extent of observed and simulated MHWs over the satellite data taking period. a, Histogram of the spatial extent of satellite-observed MHWs above the climatological (1982–2016) 99th percentile for the 1982–1998 (blue) and 1999–2016 (red) period. **b**, The spatial pattern of the MHW with the largest extent in the satellite data taking period (1982–2016), which occurred on 2 September 2015 in the North Pacific and was associated with the 'the Blob'[21]. It had a spatial extent of about twice the area of the United States (that is, 1.85×10^7 km^2). Shown are SST anomalies above the 1982–2016 climatological 99th percentile on 2 September 2015. The colour bar shows degrees Celsius. **c**, Comparison between satellite-based observations (black line) and simulations (red histogram) of the spatial extent of MHWs above the climatological 99th percentile over the 1982–2016 period. The number of MHWs (y axis) is normalized with the total number of MHWs. Deeper red colour indicates a greater number of overlapping models.



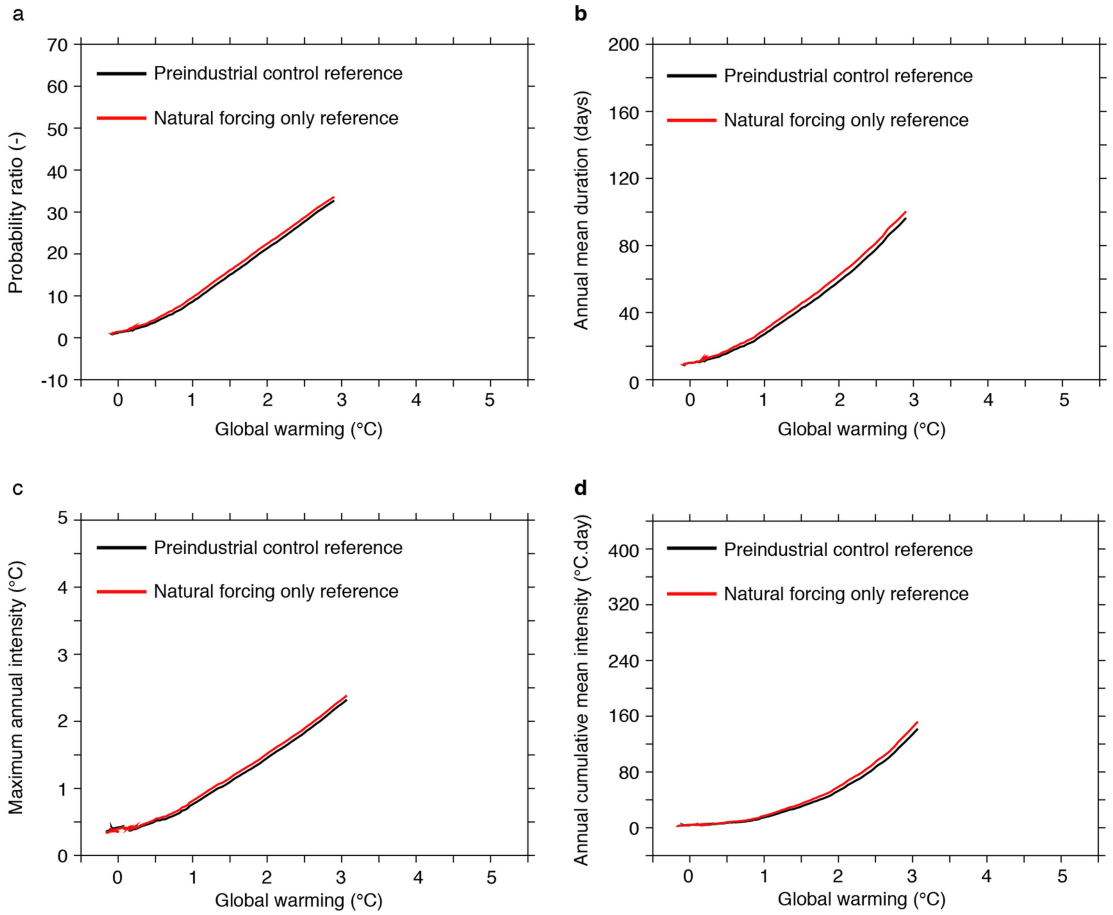
Extended Data Fig. 3 | Simulated multi-model mean changes in different MHW characteristics exceeding the preindustrial 99th percentile since preindustrial times for different global warming levels. **a**–**l**, Changes in the maximum annual intensity (**a**, **d**, **g**, **j**), annual mean duration (**b**, **e**, **h**, **k**) and annual cumulative mean intensity (**c**, **f**, **i**, **l**) of MHWs for global warming of 1 °C (**a**–**c**), 1.5 °C (**d**–**f**), 2 °C (**g**–**i**) and 3.5 °C (**j**–**l**). The black contours highlight the pattern of changes.



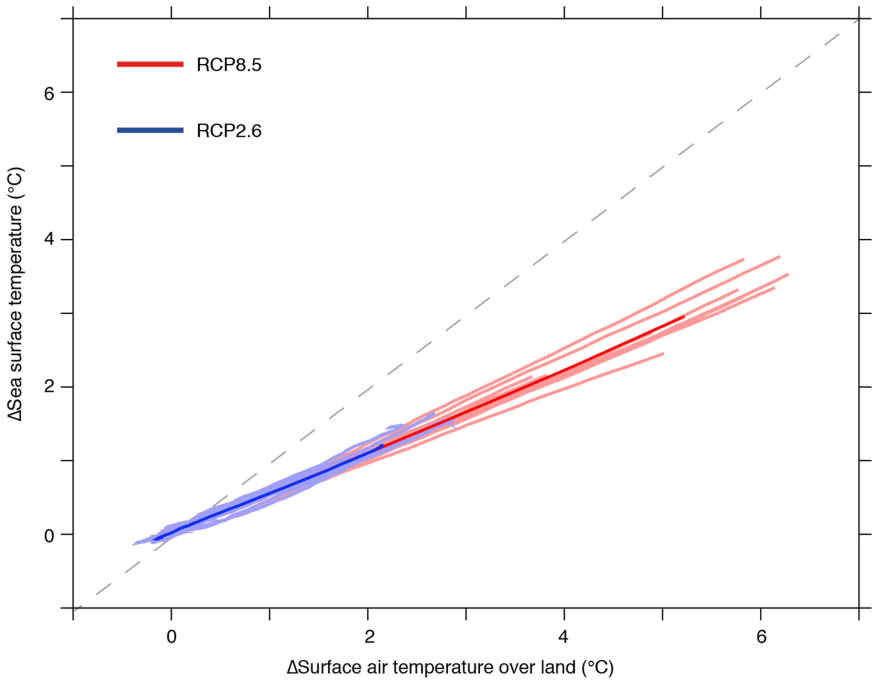
Extended Data Fig. 4 | Simulated changes in different MHW characteristics exceeding the preindustrial 99th percentile. The data were obtained from a 10-member ensemble simulation with NCAR-DOE CESM[34]. **a**–**d**, The probability ratio (**a**), annual mean duration (**b**), maximum annual intensity (**c**) and annual cumulative mean intensity (**d**) of MHWs. The black lines show the individual ensemble members. The red line in **a** shows the probability ratio versus global warming for a reference period that is defined as the 99th percentile over the 2007–2026 period, obtained using all ten ensemble members. The ensemble members are initialized from different starting points (ocean, sea ice, land and atmosphere) in the preindustrial control simulation. The simulations follow the RCP 8.5 scenario over the 21st century. The time series are smoothed with a 20-year running mean.



Extended Data Fig. 5 | Simulated changes in MHW characteristics for different global warming levels and different extreme thresholds. **a**, **c**–**e**, Global annual mean probability ratio (**a**; logarithmic scale), duration (**c**), maximum intensity (**d**) and cumulative mean intensity (**e**) of MHWs versus different extreme thresholds for different changes in global mean surface air temperature. **b**, Changes in the ratio of the mean spatial extent of MHWs between global warming conditions and 1861–1880 conditions. Simulations following only the RCP 8.5 scenario are shown. The shaded areas indicate the maximum range simulated by the CMIP5 models.



Extended Data Fig. 6 | Comparison between simulated changes in MHW characteristics exceeding the 99th percentile from a natural-forcing simulation and from a preindustrial control simulation using GFDL ESM2M forced with the RCP 8.5 scenario over the 21st century. **a**–**d**, The probability ratio (**a**), annual mean duration (**b**), maximum annual intensity (**c**) and annual cumulative mean intensity (**d**) of MHWs. The red line shows the simulated changes exceeding the 99th percentile from a natural-forcing simulation of GFDL ESM2M forced with observed solar and volcanic boundary conditions, but with greenhouse gases and aerosol concentrations set to preindustrial. The black line shows the simulated changes exceeding the preindustrial 99th percentile. The time series are smoothed with a 20-year running mean.



Extended Data Fig. 7 | Simulated changes in global SST as a function of global surface air temperature over land. The light red and blue lines represent individual model projections, whereas the dark red and blue lines represent multi-model averages for the RCP 8.5 (red) and RCP 2.6 (blue) scenarios. The time series are smoothed with a 20-year running mean. The grey dashed 1:1 line is shown for reference.

Extended Data Table 1 | Simulated changes in the annual mean spatial extent of MHWs relative to preindustrial times, and simulated annual mean duration, maximum and cumulative mean intensity of MHWs exceeding the preindustrial 99th percentile for different global warming levels

Warming	Relative change in annual mean spatial extent (-)	Annual mean duration (days)	Maximum annual intensity (°C)	Annual cumulative mean intensity (°C days)
Preindustrial	-	11 (6/14)	0.41 (0.31/0.52)	3 (2/4)
1°C	2.6 (3.4/1.3)	25 (15/33)	0.78 (0.62/1.01)	13 (8/18)
1.5°C	5.6 (11.5/3.0)	39 (24/49)	1.11 (0.83/1.40)	28 (19/42)
2°C	9.1 (18.0/6.0)	55 (36/71)	1.45 (1.09/1.79)	49 (30/71)
3.5°C	20.7 (29.1/14.9)	112 (92/129)	2.53 (2.12/2.93)	164 (126/214)

Multi-model mean estimates are shown and the associated minimum and maximum model estimates are given in parentheses.

Extended Data Table 2 | Simulated probability ratio and fraction of attributable risk estimates averaged over the ocean and over land for different global warming levels and for different preindustrial percentile thresholds (99th and 99.99th)

	99^{th} percentile		99.99^{th} percentile
Probability ratio			
Warming	Ocean	Land	Ocean
1°C	8.94 (5.70/12.15)	5.56 (3.44/8.75)	141 (47/296)
1.5°C	15.65 (11.34/24.48)	9.71 (6.63/13.48)	418 (197/1079)
2°C	22.80 (16.31/30.76)	13.89 (10.71/18.59)	893 (448/1679)
3.5°C	41.19 (36.32/44.91)	26.50 (22.36/30.26)	2560 (2094/2913)
Fraction of attributable risk			
Warming	Ocean	Land	Ocean
1°C	0.87 (0.78/0.91)	0.78 (0.65/0.87)	0.98 (0.94/0.99)
1.5°C	0.93 (0.90/0.96)	0.88 (0.83/0.92)	1 (0.99/1)
2°C	0.95 (0.94/0.97)	0.92 (0.90/0.95)	1
3.5°C	0.97 (0.97/0.98)	0.96 (0.95/0.96)	1

Values are multi-model means, and minimum and maximum model estimates are shown in parentheses.

Extended Data Table 3 | Global climate models used in this study

Model
CanESM2 (800)*
CSIRO-Mk3-6-0 (500)
GFDL-CM3 (500)
GFDL-ESM2G (500)*
GFDL-ESM2M (500)*
HadGEM2-ES (500)*
IPSL-CM5A-LR (1000)*
IPSL-CM5A-MR (300)*
MIROC-ESM (630)*
MPI-ESM-LR (1000)*
MPI-ESM-MR (150)
MRI-CGCM3 (500)

The numbers indicate the length of the control simulation analysed.
*Models used to calculate the diagnosed cumulative carbon emissions shown in Fig. 1a.

LETTER

https://doi.org/10.1038/s41586-018-0377-7

Categorical perception of colour signals in a songbird

Eleanor M. Caves[1,2], Patrick A. Green[1,2], Matthew N. Zipple[1], Susan Peters[1], Sönke Johnsen[1] & Stephen Nowicki[1]*

In many contexts, animals assess each other using signals that vary continuously across individuals and, on average, reflect variation in the quality of the signaller[1,2]. It is often assumed that signal receivers perceive and respond continuously to continuous variation in the signal[2]. Alternatively, perception and response may be discontinuous[3], owing to limitations in discrimination, categorization or both. Discrimination is the ability to tell two stimuli apart (for example, whether one can tell apart colours close to each other in hue). Categorization concerns whether stimuli are grouped based on similarities (for example, identifying colours with qualitative similarities in hue as similar even if they can be distinguished)[4]. Categorical perception is a mechanism by which perceptual systems categorize continuously varying stimuli, making specific predictions about discrimination relative to category boundaries. Here we show that female zebra finches (*Taeniopygia guttata*) categorically perceive a continuously variable assessment signal: the orange to red spectrum of male beak colour. Both predictions of categorical perception[5] were supported: females (1) categorized colour stimuli that varied along a continuum and (2) showed increased discrimination between colours from opposite sides of a category boundary compared to equally different colours from within a category. To our knowledge, this is the first demonstration of categorical perception of signal-based colouration in a bird, with implications for understanding avian colour perception and signal evolution in general.

First described for the perception of phonemes in human speech[6], categorical perception was later shown to function in the perception of auditory signals in other animals[7–9]. With regard to colour, animals may discriminate among colours but nevertheless treat them as similar[10–12], and colour categorization[13] may influence decision-making thresholds[14]. Thus, animals can categorize colours, forming discrete groups of similar yet discriminable variants across the visible spectrum. However, the hallmark of categorical perception[15,16]—increased discrimination of variants between categories relative to variants from within—has not been demonstrated for natural variation in colour-based signals.

Carotenoid-based colouration is commonly used in visual signalling across many taxa, although its function in assessment signalling is best described in mate choice in birds and fish[17,18]. Individuals vary in their ability to acquire[19] and metabolize[20] carotenoids; therefore, variation in carotenoid-based colouration has been linked to variation in the quality of the signaller. Carotenoid-based beak colouration in male zebra finches ranges from light orange to dark red[21], beak redness correlates positively with variation in cell-mediated immunity[21,22], and females show a mating preference for males with red versus orange beaks[23]. Previous studies[18] have tested how receivers respond to both ends of a carotenoid-based colour continuum, but whether they perceive variation continuously (responding differently to any detectable differences in colour) or exhibit categorical perception is unknown.

To create stimuli that vary continuously along a spectrum from red to orange, we selected eight Munsell colours (Pantone) previously used to describe the colour of zebra finch beaks[24–26]. We modelled chromatic distance (ΔS) using the receptor noise-limited model of colour discrimination[27] (Extended Data Tables 1, 2). To ensure that the selected colours are approximately equidistant from one another when accounting for zebra finch spectral sensitivity[28] and ambient light (Extended Data Figs. 1, 2 and Supplementary Information), we visualized ΔS in a chromaticity space[29] in which the Euclidean distance between stimuli plotted in an x–y plane equals chromatic distance for a trichromatic viewer (Fig. 1).

We used a food-reward protocol to test for categorization (which is sometimes referred to as 'labelling'[6]) and discrimination of the eight stimuli spanning the orange–red colour spectrum. We created discs of Munsell paper comprising two semi-circular halves of either the same or different colours (hereafter 'solid' and 'bicolour', respectively). Once females had been trained to flip over these discs to access a food reward, we trained them to flip bicolour discs first, before any solid discs; in essence, birds learned to recognize 'bicolour' versus 'solid' discs, rather than particular colour combinations. In experimental trials, we presented females with a foraging grid that had twelve wells, six of which were covered with discs: two solid discs each for each of the two colours,

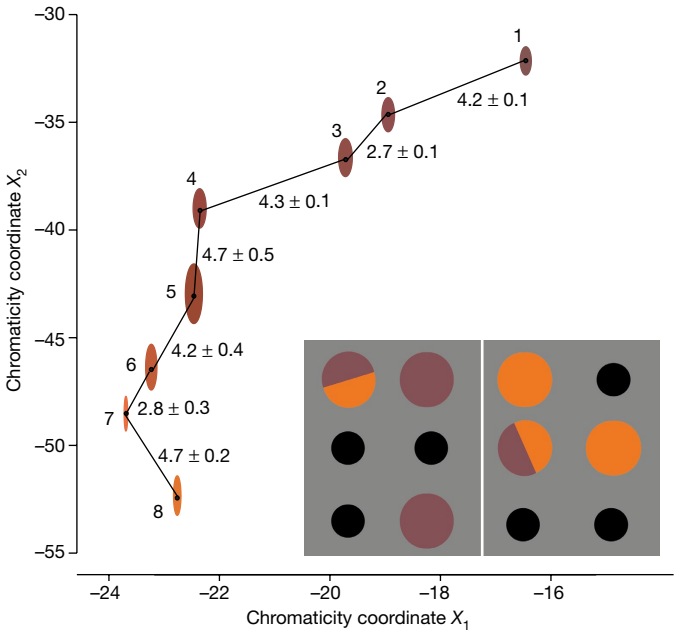


Fig. 1 | Munsell colours used to create stimuli. Colours were approximately equally spaced in chromaticity space and were closer to their nearest neighbour than to any other colour. Dots show mean chromaticity coordinates for each colour; ellipses show one standard deviation in the X_1 and X_2 dimensions ($n = 4$ measurements per colour; Supplementary Information); numbers between dots show chromatic distance (ΔS) between colours (mean ± s.d.). Ellipse colour corresponds to relevant Munsell colour (exact colours in the figure may vary). Inset, foraging grid of an example trial.

[1]Department of Biology, Duke University, Durham, NC, USA. [2]These authors contributed equally: Eleanor M. Caves, Patrick A. Green. *e-mail: snowicki@duke.edu

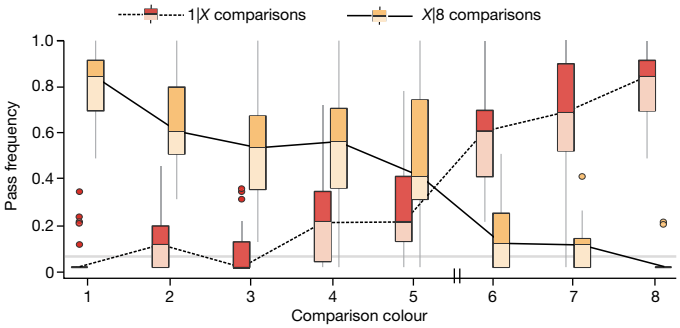


Fig. 2 | Categorization trials suggested a boundary between colours 5 and 6. The boundary is indicated by vertical lines on the *x* axis. Females were 31% more likely to pass 1|6 than 1|5 trials and 34% more likely to pass 5|8 than 6|8 trials. $n = 26$ birds in three independent cohorts. Box plots depict median (horizontal line inside box), 25th and 75th percentiles (box), 25/75th percentiles ±1.5× interquartile range (whiskers), and outliers (circles). Horizontal grey line indicates the expected pass frequency if birds flip discs by chance.

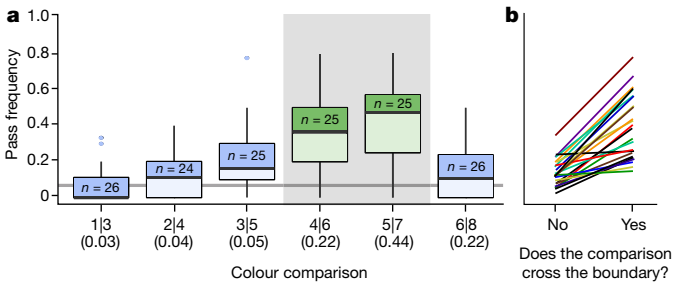


Fig. 3 | Discrimination of stimuli that are two colour spaces apart increased across the 5–6 boundary. a, **b**, Mean pass frequency across all individuals (**a**) and for each individual (**b**) was greater for comparisons that crossed the boundary versus those that did not. **a**, Green boxes in the grey-shaded area are comparisons that cross the boundary. Sample sizes (number of birds, across three independent cohorts) are shown within each box. Michelson contrast is shown in parentheses. Boxes, whiskers, circles and horizontal grey line are as described in Fig. 2.

and two bicolour discs comprising the same colours as the solid discs (Fig. 1 inset). Birds passed a trial if they flipped both bicolour discs before any solid discs, indicating that they perceived the two colours on the bicolour disc as different.

We first performed categorization experiments to establish the location of potential perceptual boundaries. We tested females using bicolour discs that included colour 1 in combination with all other colours (that is, 1|2, 1|3, 1|4 and so on) and, separately, colour 8 in combination with all other colours (that is, 8|7, 8|6, 8|5 and so on). We determined the proportion of trials passed for each comparison. For both the 1|*X* and 8|*X* comparisons (where *X* is any other colour), pass frequency increased when the chromatic distance increased between colours 1 or 8 and the comparison colour *X*. The greatest difference occurred, however, when comparing the pass frequencies for 1|5 and 1|6, and likewise between 8|5 and 8|6, suggesting that there is a potential boundary between colours 5 and 6 (Fig. 2). A linear mixed model (Table 1) demonstrated that comparing colour 1 or 8 with the colour immediately preceding versus immediately following the putative boundary resulted in the same change in pass frequency as would moving 10.5 ΔS units, approximately equal to three colour steps within a category (mean ΔS between colour steps = 3.6).

We next determined whether discrimination of colour differences crossing the putative 5–6 category boundary was increased compared to discrimination of equal colour differences that did not cross the boundary. We presented females with colour pairs that did or did not cross the hypothesized boundary, and that were two colour spaces apart (that is, 1|3, 2|4, 3|5 and so on; Fig. 3a). The pass frequency for comparisons that crossed the 5–6 boundary was 26 ± 6 percentage points higher (mean ± s.d.) than comparisons that did not (paired *t*-test, $t_{25} = 9.26$, $P < 0.0001$; Fig. 3b). We found similar results when stimuli were one or three colour spaces apart (Extended Data Figs. 3, 4) and when we combined all categorization and discrimination data into a single linear mixed model (Extended Data Table 3).

Our colour stimuli varied in brightness because real zebra finch beaks of different colour are also not of equal brightness. Several lines of evidence support the conjecture that brightness alone does not completely explain our results. First, we built a linear mixed model comparable to the one presented in Table 1, but which—in addition to chromatic distance (ΔS)—included Michelson contrast (a measure of brightness ratios) between colour pairs to explain pass frequency instead of the binary variable indicating the 5–6 boundary. This model performed substantially worse (ΔAkaike information criterion = 13) than the model that included the binary parameter of crossing the boundary. Additionally, our data show that several discrimination comparisons had either similar Michelson contrasts but different pass frequencies (4|6 compared to 6|8 (Fig. 3) and 5|6 compared to 6|7 (Extended Data Fig. 3)), or equivalent pass frequencies with different Michelson contrasts (4|6 compared to 5|7 (Fig. 3) and 3|6 compared to 4|7 and 5|8 (Extended Data Fig. 4)). Lastly, we performed one-apart discrimination experiments ($n = 18$ birds) using shades of grey (Extended Data Fig. 5 and Supplementary Information) selected to match the red–orange colours in zebra finch double cone[28] quantum catch (Extended Data Table 1), an estimate of perceived brightness in passerines[30]. In these greyscale experiments, pass frequency was significantly predicted by Michelson contrast: birds demonstrated increased discrimination for both 5|6 and 6|7 compared to greyscale pairs with a lower contrast (Table 1 and Extended Data Fig. 5). Together, these pieces of evidence suggest that brightness may contribute to category formation but cannot alone explain the perceptual categories that we observed.

Our results demonstrate categorical perception of colour associated with an assessment signal. We found that a category boundary influences the perception of two colour stimuli as similar or different, and results in differential discrimination between stimuli depending on their location relative to the category boundary. Discrimination improved with increasing chromatic distance between colours (that is, variants within a category were not perceived as identical), but increased most sharply across the category boundary (that is, variants from across the boundary were perceived as most different).

We did not explicitly test whether categorical perception arises at the level of the photoreceptor or higher, such as in the brain. However, a wavelength discrimination function derived from electroretinographic data from the pigeon *Columba livia*[31], a reasonable proxy for the zebra

Table 1 | Mixed model results for categorization and greyscale discrimination data

Model	Parameter	Estimate	χ^2 (d.f. = 1)	*P* value
Colour categorization	Intercept	0.011		
	$\Delta S = 1$	0.022	77.8	<0.0001
	Across 5–6 boundary	0.24	30.8	<0.0001
Greyscale discrimination	Intercept	0.001		
	Michelson contrast	2.30	106	<0.0001

A linear mixed model built from categorization data demonstrates significant effects on pass rate of chromatic distance between colours and whether the comparison crossed the 5–6 boundary. Bird ID was included as a random effect. The binary parameter of crossing the boundary was collinear with differences in brightness; therefore, both could not be included in the same model. Including a random slope of whether the comparison crossed the 5–6 boundary significantly improved the model ($\chi^2(2) = 28.6$, $P < 0.0001$), indicating significant inter-bird variation in the magnitude of the effect of crossing the 5–6 boundary. This variation may be due to variation in boundary location between birds and/or variation in the strength of the effect of a comparison crossing a bird's categorical boundary. A model that also included a random slope of perceptual distance did not improve the model (likelihood ratio test, $P > 0.1$) or qualitatively change the results. A mixed effects model built from greyscale discrimination demonstrates a significant effect of Michelson contrast on pass rate in the absence of hue information. Bird ID was included as a random effect. *P* values were calculated based on ANOVA comparisons of models that do or do not include each parameter.

finch, suggests that photoreceptor sensitivity alone probably does not explain categorical perception (Extended Data Fig. 6). We also did not directly test female preference for males in relation to the categories found here. Nonetheless, these findings have important implications for the understanding of colour perception and encourage further work exploring whether and how categorical perception interacts with selection on signal form and function, particularly in the context of assessment (Extended Data Fig. 7).

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0377-7.

Received: 20 November 2017; Accepted: 21 June 2018;
Published online 1 August 2018.

1. Maynard Smith, J. & Harper, D. *Animal Signals* (Oxford Univ. Press, Oxford, 2003).
2. Searcy, W. A. & Nowicki, S. *The Evolution of Animal Communication* (Princeton Univ. Press, Princeton, 2005).
3. Roff, D. A. The evolution of mate choice: a dialogue between theory and experiment. *Ann. NY Acad. Sci.* **1360**, 1–15 (2015).
4. Bornstein, M. H. in *Categorical Perception: the Groundwork of Cognition* (ed. Harnad, S. R.) 287–300 (Cambridge Univ. Press, Cambridge, 1987).
5. Studdert-Kennedy, M., Liberman, A. M., Harris, K. S. & Cooper, F. S. Motor theory of speech perception: a reply to Lane's critical review. *Psychol. Rev.* **77**, 234–249 (1970).
6. Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* **54**, 358–368 (1957).
7. Baugh, A. T., Akre, K. L. & Ryan, M. J. Categorical perception of a natural, multivariate signal: mating call recognition in túngara frogs. *Proc. Natl Acad. Sci. USA* **105**, 8985–8988 (2008).
8. Wyttenbach, R. A., May, M. L. & Hoy, R. R. Categorical perception of sound frequency by crickets. *Science* **273**, 1542–1544 (1996).
9. Nelson, D. A. & Marler, P. Categorical perception of a natural stimulus continuum: birdsong. *Science* **244**, 976–978 (1989).
10. Wright, A. A. & Cumming, W. W. Color-naming functions for the pigeon. *J. Exp. Anal. Behav.* **15**, 7–17 (1971).
11. von Frisch, K. *Bees: Their Vision, Chemical Senses, and Language* (Cornell Univ. Press, Ithaca, 1964).
12. Sandell, J. H., Gross, C. G. & Bornstein, M. H. Color categories in macaques. *J. Comp. Physiol. Psychol.* **93**, 626–635 (1979).
13. Jones, C. D., Osorio, D. & Baddeley, R. J. Colour categorization by domestic chicks. *Proc. R. Soc. B* **268**, 2077–2084 (2001).
14. Hanley, D. et al. Egg discrimination along a gradient of natural variation in eggshell coloration. *Proc. R. Soc. B* **284**, 20162592 (2017).
15. Harnad, S. R. *Categorical Perception: the Groundwork of Cognition* (Cambridge Univ. Press, Cambridge, 1987).
16. Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L. & Rowland, D. Categorical perception of morphed facial expressions. *Vis. Cogn.* **3**, 81–118 (1996).
17. Endler, J. A. Natural and sexual selection on color patterns in poeciliid fishes. *Environ. Biol. Fishes* **9**, 173–190 (1983).
18. Hill, G. E. Female house finches prefer colourful males: sexual selection for a condition-dependent trait. *Anim. Behav.* **40**, 563–572 (1990).
19. Hill, G. E., Montgomerie, R., Inouye, C. Y. & Dale, J. Influence of dietary carotenoids on plasma and plumage colour in the house finch: intra- and intersexual variation. *Funct. Ecol.* **8**, 343–350 (1994).
20. Borel, P. Genetic variations involved in interindividual variability in carotenoid status. *Mol. Nutr. Food Res.* **56**, 228–240 (2012).
21. Blount, J. D., Metcalfe, N. B., Birkhead, T. R. & Surai, P. F. Carotenoid modulation of immune function and sexual attractiveness in zebra finches. *Science* **300**, 125–127 (2003).
22. George, D. B., Schneider, B. C., McGraw, K. J. & Ardia, D. R. Carotenoids buffer the acute phase response on fever, sickness behavior and rapid bill color change in zebra finches. *J. Exp. Biol.* **220**, 2957–2964 (2017).
23. Collins, S. A. & ten Cate, C. Does beak colour affect female preference in zebra finches? *Anim. Behav.* **52**, 105–112 (1996).
24. Burley, N. & Coopersmith, C. B. Bill color preferences of zebra finches. *Ethology* **76**, 133–151 (1987).
25. Collins, S. A., Hubbard, C. & Houtman, A. M. Female mate choice in the zebra finch — the effect of male beak colour and male song. *Behav. Ecol. Sociobiol.* **35**, 21–25 (1994).
26. Birkhead, T. R., Fletcher, F. & Pellatt, E. J. Sexual selection in the zebra finch *Taeniopygia guttata*: condition, sex traits and immune capacity. *Behav. Ecol. Sociobiol.* **44**, 179–191 (1998).
27. Vorobyev, M. & Osorio, D. Receptor noise as a determinant of colour thresholds. *Proc. R. Soc. B* **265**, 351–358 (1998).
28. Lind, O. Colour vision and background adaptation in a passerine bird, the zebra finch (*Taeniopygia guttata*). *R. Soc. Open Sci.* **3**, 160383 (2016).
29. Hempel de Ibarra, N., Giurfa, M. & Vorobyev, M. Discrimination of coloured patterns by honeybees through chromatic and achromatic cues. *J. Comp. Physiol. A* **188**, 503–512 (2002).
30. Osorio, D. & Vorobyev, M. Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. *Proc. R. Soc. B* **272**, 1745–1752 (2005).
31. Riggs, L. A., Blough, P. M. & Schafer, K. L. Electrical responses of the pigeon eye to changes in wavelength of the stimulating light. *Vision Res.* **12**, 981–991 (1972).

Acknowledgements We thank D. Osorio for comments on the manuscript; L. Jin, K. Kochvar, J. Bollinger and L. Deehan for assistance running trials; and R. Mooney and M. Booze for providing access to zebra finches from their colony. Funding for this study was provided by the Duke University Office of the Provost.

Reviewer information *Nature* thanks A. Kelber and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.N. conceived and led the project, and secured funding; E.M.C., P.A.G. and M.N.Z. designed and carried out behavioural experiments, with assistance from S.P.; S.J. led the visual modelling and colour vision aspects of the study, with assistance from E.M.C. and P.A.G.; M.N.Z. conducted statistical analyses; P.A.G. prepared figures with assistance from M.N.Z.; S.P. managed the zebra finch population; E.M.C. wrote the initial draft of the manuscript, which was subsequently edited and approved by all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0377-7.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0377-7.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to S.N.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

The goal of this study was to test whether female zebra finches perceive colours along a red-to-orange spectrum in a continuous or categorical fashion. We selected eight colours from red to orange that correspond to male zebra finch beak colours and that are roughly equidistant from one another in a chromaticity space based on zebra finch spectral sensitivity. These eight colours were made into disc stimuli that were either solid in colour (that is, made of two halves of the same colour) or bicolour (that is, one colour on one half of the disc and another colour on the other half). Disc stimuli were used in a food-reward protocol. Female zebra finches were first trained to flip over discs covering wells in a foraging grid to find a food reward (millet seeds, as used previously[32]), using discs made from the colours at the endpoints of our red–orange continuum (colours 1 and 8). Initially, all wells (both solid and bicolour) were baited with millet seeds, to reward the birds any time they successfully flipped a disc. After being trained on this task, females were further trained to flip only bicolour discs; we used the same stimuli made of colours 1 and 8, but this time baited only wells that were covered with bicolour discs. Once females passed six out of seven consecutive trials under this protocol, indicating they had learned to search for food only under bicoloured discs, we then conducted trials using different combinations of our eight selected colours. If females perceived two colours as distinct, they would flip the bicoloured discs first to gain a food reward; if they did not perceive two colours as distinct, they would not preferentially flip the bicoloured discs first. For details, see 'Behavioural trials'.

Birds used in this study. All birds in this study were sexually mature female zebra finches (age range: 3–50 months at start of experimental testing) from a colony maintained by R. Mooney at Duke University (IACUC A258-14-10). After transfer from the colony, birds were housed singly in cages ($46 \times 23 \times 23\,cm^3$, Prevue Pet) outfitted with two wooden perches and a cuttlebone. Lighting was controlled during trials (see 'Behavioural trials') and food was removed 5 h before trials to ensure that birds would be motivated to perform the task. Outside of trial times, birds were kept on a 15 h:9 h light:dark cycle (consistent with the light cycle of the birds' home colony), with overhead lighting provided by fluorescent bulbs (Ecolux with Starcoat SP 35/41, colour temperature 3,500–4,100 K, General Electric) with ballast (Hi-Lume 3D/Eco-10, Lutron Electronics) operating at 50–60 Hz. Birds were given zebra finch food (Kaytee Forti-Diet Pro Health Finch diet) and water ad libitum. Rooms were maintained at 25–27 °C. Testing was done under Duke University IACUC protocol A004-17-01.

Selecting stimulus colours. Previous work has shown that the range of beak colouration in zebra finch males can be represented by red and orange shades in the Munsell colour system[24–26]. Munsell colours are defined by three parameters: hue, value (brightness) and chroma (saturation). Previous work on zebra finches identified a large set of Munsell colours that, by the human eye, approximate the colours of zebra finch beaks. In particular, this set of Munsell colours consists of colours with hues 6.25R–3.75YR, values 3.5–6 and chromas 10–14[24,26]. Notably, the values used in these previous studies spanned a range of hues from yellow–orange to red, and specifically did not use Munsell colours of the same brightness because real male beaks of different hues are also of different brightness.

These Munsell colours[24–26] are based on beak colours across two different colony populations of zebra finches and capture most of the variation within those populations, although the authors of the studies note that the beaks of occasional individuals were found to be outside that range. Because the goal of our study was to examine how female zebra finches perceive a range of colours that are similar to the range spanned by the colours of male zebra finch beaks, we used this previous work as a starting point to choose 40 Munsell colour samples from within the set outlined above. Reflectance spectra from each colour sample were measured using an integrating sphere with a built-in tungsten-halogen light source (ISP-REF; Ocean Optics). All measurements were taken with reference to a Spectralon 99% white reflectance standard (Labsphere).

For each of the Munsell colours, we calculated relative photon catches (that is, how many photons are absorbed by a given cone type in response to a visual scene[33]) for zebra finch short-, medium- and long-wave cones. Photon catches were calculated over the wavelengths 400−700 nm, using zebra finch spectral sensitivity curves[28], an ambient light spectrum and the reflectance spectrum of each colour. Thus, photon catch Q for receptor type r in response to colour c was calculated using:

$$Q_{r,c}(\lambda) \propto \int_{400}^{700} S_r(\lambda) * R_c(\lambda) * I(\lambda) \mathrm{d}\lambda$$

where S_r is the sensitivity of receptor type r, R_c is the reflectance of colour c, λ denotes the wavelength, and I is the irradiance of the illuminant. We use proportionality throughout, because we did not require an absolute quantum catch, and the constant factors remain the same across different receptor types. As for our measure of ambient light, we used a standard tungsten bulb illuminance spectrum (CIE Illuminant A, colour temperature 2,856 K; spectrum in Supplementary Information), which is very similar to the spectrum of ambient light provided by the tungsten bulbs in our experiment (see below, Extended Data Fig. 1 and Supplementary Information). We used sensitivity data from the zebra finch[28] to calculate photon catches for the short-, medium- and long-wave and double cones (Extended Data Table 1). We did not calculate photon catches for the ultraviolet cone because (1) male zebra finch beaks reflect minimal ultraviolet light[34], (2) female zebra finches do not require ultraviolet radiation to discriminate and rank red–orange colouration[35], (3) our own measurements confirmed that Munsell paper does not reflect strongly in the ultraviolet portion of the spectrum, and (4) the reflected ultraviolet radiance from our Munsell chip stimuli, under experimental lighting conditions, was essentially zero (see Extended Data Fig. 2).

In tests of categorical perception, stimuli should be roughly equidistant from one another in perceptual space. No metric can fully describe the perceptual space of human colour vision, let alone that of a zebra finch. Therefore, to calculate the chromatic distance between colour stimuli, we used the receptor noise-limited (RNL) model of colour discrimination[27], which uses photon catches calculated from the spectral sensitivity of a relevant visual system to calculate ΔS, a measure of chromatic distance between two colours (equivalent to just-noticeable differences, JNDs). We visualized ΔS using a perceptually uniform, two-dimensional space based on both hue and saturation/chroma, in which the Euclidean distance between two colours is equivalent to the RNL model-derived chromatic distance (ΔS) between colours[29]. Plotting chromatic distances in this two-dimensional space is only relevant for trichromatic vision, which in our case was appropriate given that we did not incorporate quantum catches from ultraviolet cones (see above). The RNL model[27] states that colour discriminability is primarily limited by photoreceptor noise; thus a Euclidean distance of 1.0 in the chromaticity space is approximately equal to one standard deviation of receptor noise, or one JND. Given that this method uses photon catches from single cones (short-, medium- and long-wave), we performed a separate analysis to examine perceived brightness based on double cones (see below).

In summary, the eight colours that we used (see Supplementary Information for reflectance spectra) were based on previously published comparisons with male zebra finch beaks and are approximately equally spaced in a chromaticity space based on zebra finch spectral sensitivity. Throughout, we refer to these colours as 1 (the darkest, red end of the range) through 8 (the brightest, orange end of the range).

Sensitivity analyses. To ensure that the eight colours that we selected are not equidistant in chromaticity space simply owing to the exact parameters of the zebra finch spectral sensitivities and the tungsten lighting conditions that we used, we additionally plotted the selected colours in chromaticity space using different conditions (all spectra and sensitivity curves are provided in the Supplementary Information), including: the zebra finch ultraviolet-light-sensitive (UVS) cone type retina[28] under both (1) tungsten and (2) daylight (D65) illuminants; (3) the zebra finch UVS cone type retina using the spectrum of halogen light present in the experimental rooms; (4) the zebra finch UVS cone type retina after applying von Kries adaptation based on a neutral grey background[36] to account for colour constancy mechanisms; (5) another UVS cone-type retina visual system (the starling *Sturnus vulgaris*[37]); (6) the average UVS cone-type retina (data from previously published supplementary online material[33]); and (7) the average violet-light-sensitive (VS) cone-type retina (data from previously published supplementary online material[33]), which is the other primary type of retina found in birds (Extended Data Table 2). Overall, chromatic distances between selected Munsell colours remained relatively consistent even when changing the spectral sensitivity and lighting conditions, as well as when accounting for colour constancy, with the distance between 5 and 6 never being the largest jump. Thus, chromatic distance alone cannot explain the category boundary that we observed.

Behavioural trials. *Room set-up.* Birds were housed in individual cages, with up to 12 birds in each room. Before trials, we placed opaque barriers between adjacent cages so that birds could not see their closest neighbours perform any tasks. Individuals were able to see other birds across the room, although they could not see the task that birds were performing—this set-up prevented birds from seeing how other birds were passing the task while ensuring that they could see other birds, which was beneficial given that zebra finches are highly social. During trials, the room's overhead lights were turned off and birds were allowed approximately 10 min to acclimatize to the experimental lighting. During experimental trials, each cage was illuminated from above by a halogen lamp (colour temperature 2,900 K, model number H&PC-61361, Philips Lighting) approximately 81 cm from the foraging grid (spectra in Supplementary Information). The light passed through vellum paper hung 8 cm from the light source to ensure that each cage had even and diffuse illumination (Extended Data Fig. 1). All trials were recorded on video (Logitech Webcam Pro 9000, Logitech). On trial days, food was removed from the cages at 09:00 and trials began at 14:00, to ensure that birds would be motivated to attempt the task.

Disc design. Our experimental stimuli were discs 2.5 cm in diameter, made of two semi-circles of Munsell colour sheets, glued with their straight edges together to create a full circle. The discs were covered with a clear epoxy cover. A clear vinyl disc (1.3 cm diameter, 0.3 cm high) was attached to the bottom of the disc to ensure it fit precisely into the wells of the foraging grid. We created two types of discs: solid, in which both semi-circular halves were the same colour; and bicolour, in which the two semi-circular halves were made of different colours. Bicolour discs were named for the colours of their two halves—that is, a disc made up of half colour 1 (the far red) and half colour 8 (the far orange) would be referred to as 1|8.

The experimental foraging grid consisted of two grey plastic blocks ($13.5 \times 9\,cm^2$ and 2.5 cm high) positioned adjacent to one another. Each block contained six identical wells (1.3 cm diameter, 0.8 cm deep). Birds first learned to search for food beneath the discs in five stages. In the first stage, we placed millet seeds in four randomly selected wells, with no discs present. In the second stage, four discs (two bicolour, and one each of the two solid colours that comprised the bicolour discs halves) were on the grid, adjacent to the baited wells. In the third stage, discs were placed halfway covering the baited wells. In stage four, discs were placed tipped into the baited wells, so as to hide the seed but with discs fitted only loosely into the well. In the fifth and final stage, discs were placed to completely cover the baited well, so a bird could only access the seed by flipping the disc off of the well using its beak. For each of the training tasks, success was defined as obtaining the seeds from any of the baited wells within a 20-min period. A subject had to pass three consecutive trials of each stage to progress to the next stage.

Once the birds had learned the basic task of searching for food under the discs, they were trained to associate only bicolour discs with a food reward, a stage that we call 'bicolour association'. Zebra finches were trained in bicolour association using a total of six discs: two 1|8 bicolour discs (which were baited), and two solid discs each of colour 1 and 8 (which were not baited). Importantly, our behavioural data show that birds learned to recognize bicolour versus solid discs, rather than particular colour combinations of bicolour discs, as shown by their ability to extrapolate the colour-based task to a greyscale task (described below). Trials lasted for 2 min, during which time birds were allowed to flip discs, and we recorded the order of the first two discs that they flipped. Following the 2-min observation period, the grid and any remaining millet were left in for up to 20 min. To pass a bicolour association trial, birds had to flip over both bicolour (1|8) discs before either solid disc was flipped within the 2-min observation period. Birds had to pass six out of seven consecutive trials before we determined that they had learned the task, after which they progressed to data collection. If an individual failed more than one training trial out of seven, they were given additional training trials until the pass criterion was reached or until we determined the bird was unlikely to learn the task. In total, 26 out of 30 birds (87%) that reached the bicolour association trials passed and went on to data collection trials.

Data collection trials involved the same grid and 12-well set up as in training. Six wells were covered with discs (two bicolour, four solid), and only the two bicolour discs were baited. We used the 'sample' function in R[38] to create a set of randomized locations for each disc in each trial, up to 6 trials for each colour combination. Discs were placed in haphazard orientations (that is, the direction of the line bisecting the middle was not consistent from one trial to the next). During data collection, birds were allowed up to 2 min to flip discs, after which time the grid and any remaining seed were removed. We recorded both the order in which the first two discs were flipped, as well as the latency to flipping the first disc. Observers were not blind to the tasks; however, up to three of five (usually, one or two) observers collected data on a given day and no single observer consistently collected data for a given trial (for example, 5|6, 1|6). Results were consistent across the five observers who collected data during the experiment.

We conducted 3–5 trials per bird per day. Overall, we collected data from 26 birds (trained in three independent cohorts of $n = 10$, $n = 6$ and $n = 10$) over the course of 7,015 experimental trials. In 10.1% of trials ($n = 712$), birds flipped either no discs or one disc and then stopped flipping discs. In these cases, we did not count these trials as a failure, but rather removed the trial from the dataset. Most commonly, birds did not flip any discs or only flipped one disc if other birds in the room suddenly went quiet or if they were startled by the behaviour of another bird. Removing these incomplete trials from the dataset ensured that we did not bias our data towards an increased failure rate. At the end of each trial day, we performed a motivation check on each bird by returning the bird's food dish to the cage and observing for up to 10 min to ensure that they would eat seeds. Birds typically began eating within 30 s of having their food returned, indicating that individuals still had high feeding motivation even at the end of our trials.

Categorization experiments. Categorization experiments tested whether birds responded in a continuous or discontinuous manner to colour variation, identifying the location of boundaries if they were present. As in all trials, we presented birds with two bicolour discs and two of each solid colour disc (six discs total spread randomly across 12 possible wells on the grid). Each day, a bird was given one refresher trial (1|8 bicolour and 1 and 8 solid) and five experimental trials of a single colour combination. The next day, the bird would receive a refresher trial and five trials on a different colour combination, and so on. Colour combinations used in categorization trials included the combination of colour one and all other colours (that is, 1|2, 1|3, 1|4…) as well as the combination of colour eight and all other colours (that is, 8|7, 8|6, 8|5…). Additionally, as a control for the possibility that birds used olfactory cues to locate seeds under certain discs, we performed 1|1 and 8|8 trials to check that, in the absence of any variation in visual information, birds would not perform better than chance at flipping discs placed on baited well. Birds were shown colour combinations in an order that alternated relatively distant colour combinations (those that were five to seven steps apart) with relatively closer colour combinations (those that were one to four steps apart). Each bird was given two experimental days (10 total trials) with each colour combination, and one experimental day (5 total trials) with each of the 1|1 and 8|8 comparisons. The second experimental day for each combination occurred only after the first experimental day for all other combinations had occurred (that is, experimental days for a specific colour combination did not immediately repeat).

Preliminary data analysis identified a putative boundary between colours 5 and 6. To analyse these data, we calculated the proportion of trials that each bird passed for each comparison (15 total comparisons per bird). Using only data from categorization trials, we built a linear mixed model (throughout, linear mixed models were calculated using R package lme4[39]) that included pass rate as its response variable, Euclidian chromaticity distance (ΔS) between the two colours being compared and whether the comparison crossed the putative 5–6 boundary as fixed effects, and bird ID as a random effect. The model included random slopes in addition to random intercepts, to account for variation among birds in the strength of the effect of crossing the 5–6 boundary.

We visually inspected the residuals of this model using a quantile–quantile plot, a histogram and a scatterplot of residuals versus predicted values to confirm that the residuals of the model were approximately normally distributed around zero and that they were homoscedastic.

Discrimination experiments. A second requirement for demonstrating categorical perception is that subjects show increased discrimination between stimuli that cross a category boundary compared to equally spaced stimuli that do not cross a boundary. Thus, for discrimination trials, bicoloured discs were made of colour pairs that were one space apart (that is, 1|2, 2|3, 3|4…; Extended Data Fig. 3), two spaces apart (that is, 1|3, 2|4, 3|5…; Fig. 3) or three spaces apart (that is, 1|4, 2|5, 3|6…; Extended Data Fig. 4). The experimental set-up and criteria for passing were the same as for categorization trials. For each bird, mean pass rates were calculated for all discrimination trials that did not cross the putative boundary and, separately, for those that did. We then calculated the difference between these means for each bird and used a two-sided *t*-test to determine whether the mean difference in pass rate was significantly different from zero.

Combining categorization and discrimination data in a single statistical model. To present our data in the most easily interpreted format, we present separate analyses for categorization and discrimination trials in the main text. To bolster our conclusions, maximize statistical power and contain our analysis in a single model, we built a linear mixed model that includes all discrimination and categorization data. As reported in the main text, this model contains random effects of bird ID and includes crossing the 5–6 boundary as a random slope. As in the main text, significance tests were performed using ANOVA comparisons. The results of the single model can be found in Extended Data Table 3.

Greyscale discrimination experiments. The set of eight Munsell colours that we used in this experiment were selected primarily based on their colour: in particular, the colour of the Munsell colours aligned with those of actual beaks and were equally spaced in a chromaticity space that describes the hue and saturation of colours and from which brightness has been factored out. However, brightness is also an important part of how colour is perceived. To examine the effects that brightness has on structuring zebra finch perception of colours, we performed a second experiment in which we used shades of grey that matched our eight Munsell colours in brightness, as perceived by zebra finches.

To select appropriate shades of grey, we used an integrating sphere to measure reflectance spectra from a set of 72 grey paint swatches (Behr brand, Behr Process). We then calculated the quantum catch of the zebra finch double cone[28] for each of our eight Munsell colours and for each of the grey paint swatches; the double cone is believed to be the primary way in which birds encode brightness information[30]. By finding shades of grey that matched each Munsell colour in double cone quantum catch, we selected a set of eight grey shades that were equivalent in bird-perceived brightness to the colour stimuli that we had been using.

Using this set of eight shades of grey, we created experimental discs for the 1|8-grey and each one-apart combination of grey shades. We followed the same experimental procedure as above to examine the birds' discrimination ability when hue information had been removed (Extended Data Fig. 5).

Wavelength discrimination function. To investigate whether the categorical perception that we observed could be solely due to the wavelength discrimination function (WDF) of avian photoreceptors, we examined a WDF that was derived using electroretinography in the pigeon *C. livia*[31]. We know of no experimentally or behaviourally derived WDF for zebra finches. The pigeon is a reasonable substitute, however, given that the spectral sensitivity peaks for its medium- and long-wavelength sensitive cones (505 and 565 nm, respectively[40]) are very similar to those of the zebra finch (507 and 565 nm, respectively[28]).

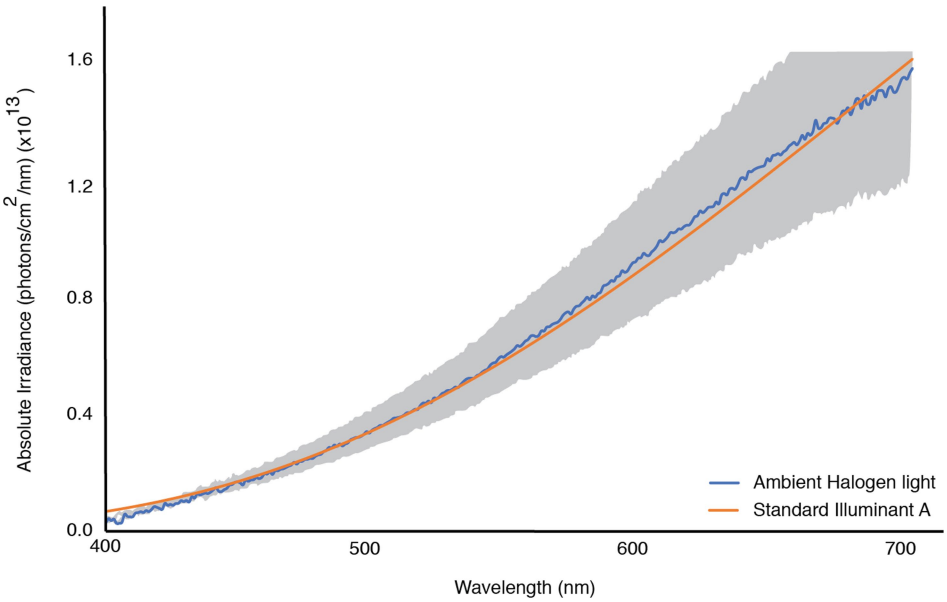
We plotted electroretinography-derived WDF data[31] alongside the spectral sensitivity peaks of the zebra finches (Extended Data Fig. 6). The units of the WDF are arbitrary units, in keeping with the original publication, but show the general pattern of wavelength discrimination. One complication is that it is not possible to know precisely where in the visible spectrum the avian perception of 'orange' and 'red' would occur. However, in humans, orange and red both occur above 580 nm, in the region in which stimulation is primarily of the long-wavelength cone and secondarily the medium-wavelength cone. The WDF showed a discriminability peak at approximately 605 nm, followed by a relatively smooth decrease in discriminability in the region of the spectrum in which avian viewers probably see orange and red (Extended Data Fig. 6). The shape of these curves suggests that photoreceptor sensitivity on its own probably does not explain categorical perception. From this analysis alone, however, we cannot rule out that hue discrimination based on spectral sensitivity curves contributes to categorical perception. Ultimately, whether categorical perception arises at the level of the photoreceptor or retina, or is a higher-order process, does not affect our findings or interpretation, but suggests avenues for future research.

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

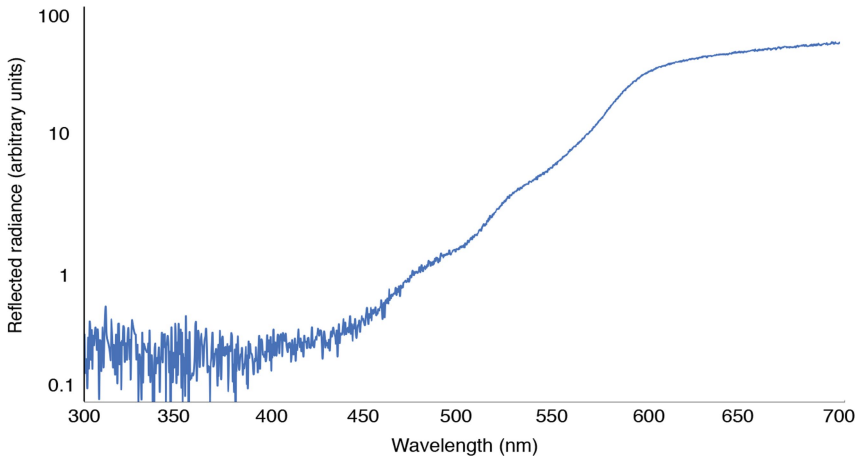
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The datasets generated and analysed during the current study are available in the Duke Data Repository: https://doi.org/10.7924/r4rx96r99.

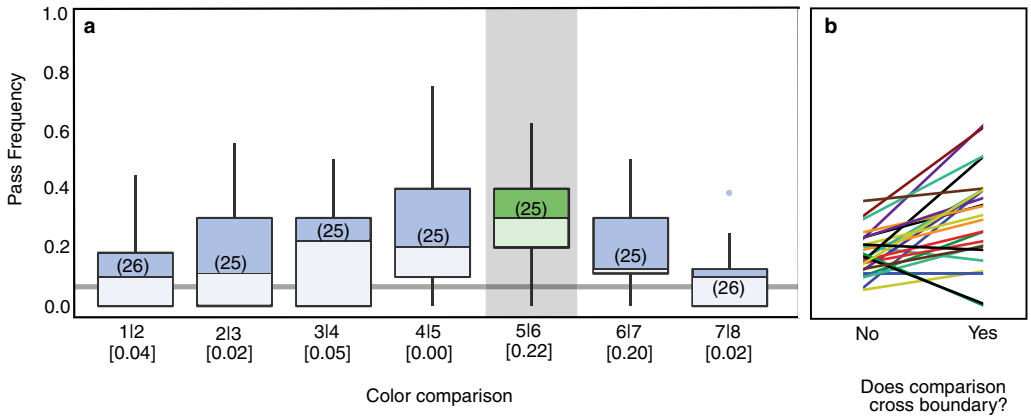
32. Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A. & Nowicki, S. Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Anim. Behav.* **81**, 1209–1216 (2011).
33. Endler, J. A. & Mielke, P. W. Comparing entire colour patterns as birds see them. *Biol. J. Linn. Soc.* **86**, 405–431 (2005).
34. Bennett, A. T. D., Cuthill, I. C., Partridge, J. C. & Maier, E. J. Ultraviolet vision and mate choice in zebra finches. *Nature* **380**, 433–435 (1996).
35. Hunt, S., Cuthill, I. C., Swaddle, J. P. & Bennett, A. T. D. Ultraviolet vision and band-colour preferences in female zebra finches, *Taeniopygia guttata*. *Anim. Behav.* **54**, 1383–1392 (1997).
36. Balkenius, A. & Kelber, A. Colour constancy in diurnal and nocturnal hawkmoths. *J. Exp. Biol.* **207**, 3307–3316 (2004).
37. Stoddard, M. C. & Prum, R. O. Evolution of avian plumage color in a tetrahedral color space: a phylogenetic analysis of new world buntings. *Am. Nat.* **171**, 755–776 (2008).
38. R Core Team. *R: A Language and Environment for Statistical Computing* http://www.R-project.org/ (R Foundation for Statistical Computing, Vienna, 2016).
39. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
40. Vorobyev, M., Osorio, D., Bennett, A. T. D., Marshall, N. J. & Cuthill, I. C. Tetrachromacy, oil droplets and bird plumage colours. *J. Comp. Physiol. A* **183**, 621–633 (1998).



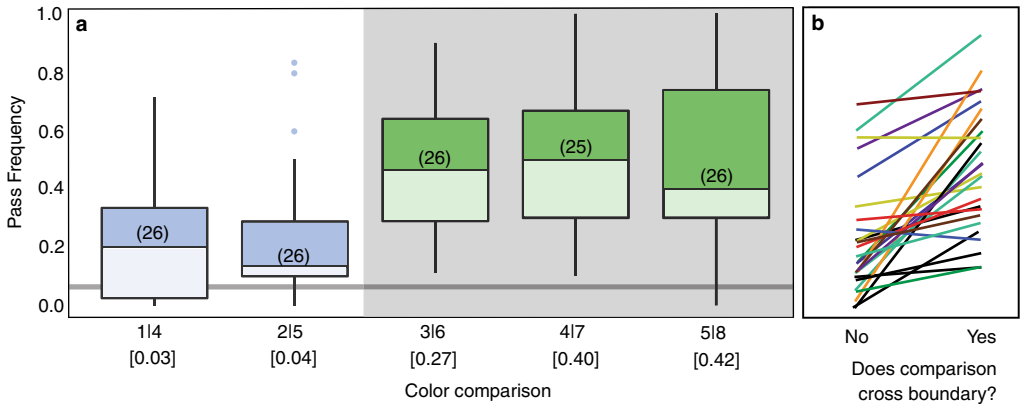
Extended Data Fig. 1 | Downwelling vector irradiance at the level of the foraging grid. Units are photons per cm^2 per nm. Tungsten bulbs were used to illuminate each cage from a set distance. There was some variation in irradiance between cages. The blue line represents the mean absolute irradiance of our halogen bulbs, and the grey-shaded region indicates one standard deviation in either direction. The orange line illustrates standard Illuminant A, which we used throughout for visual modelling because it is a standard spectrum (and thus repeatable by other researchers) and closely matched the ambient lighting in our room.



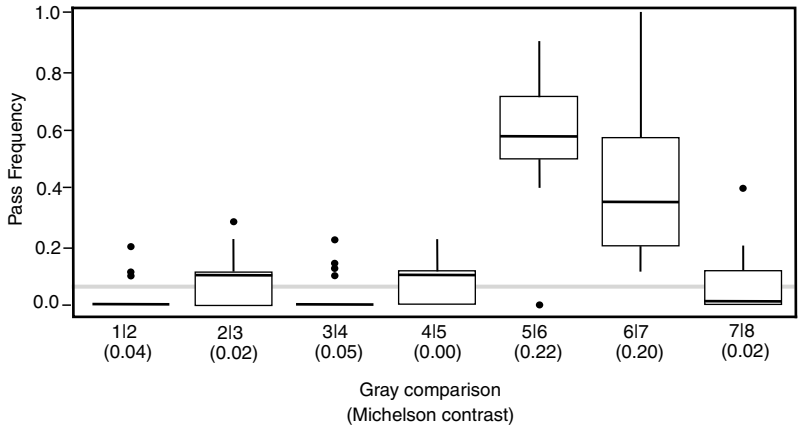
Extended Data Fig. 2 | Reflected radiance of experimental stimuli under experimental lighting conditions. Reflected radiance is in arbitrary energy units; stimulus is Munsell paper covered with an epoxy cover. Below 400 nm, the values are so low that they reach the noise floor of the spectrometer, suggesting that the use of a trichromatic model (versus one that includes ultraviolet sensitivity) to examine chromatic distance between colours is appropriate.



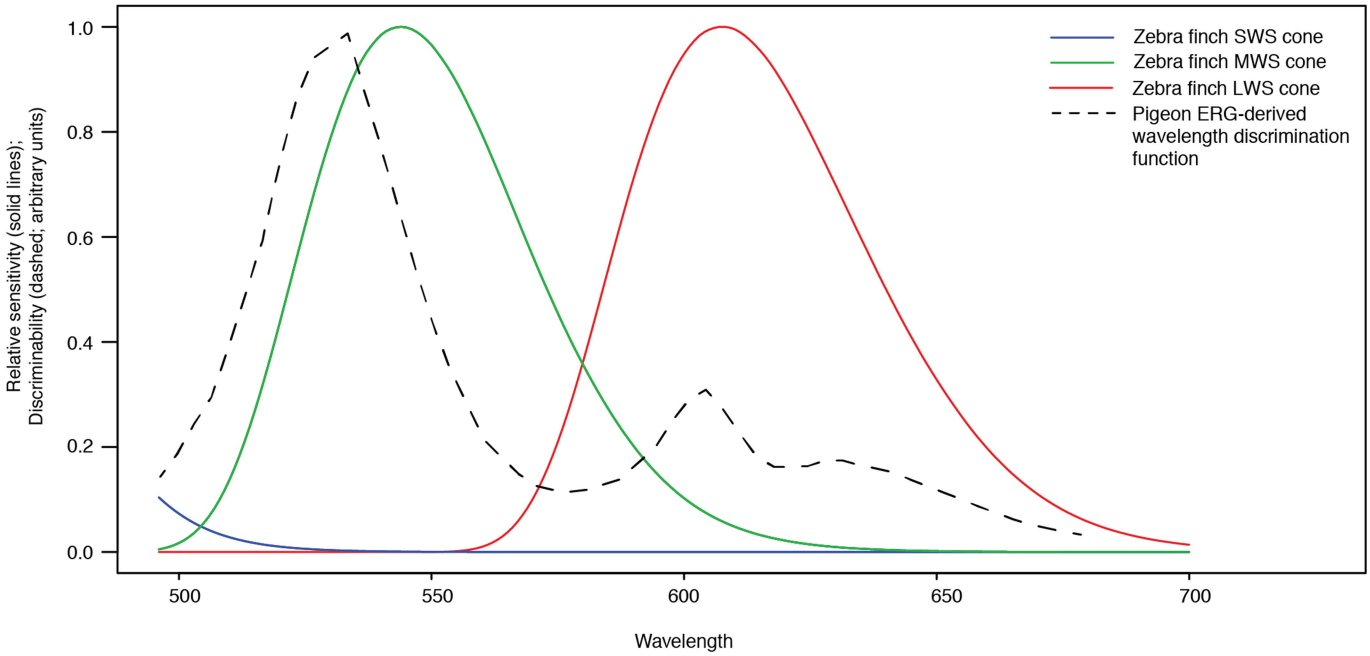
Extended Data Fig. 3 | Results from one-apart discrimination trials. **a**, **b**, Results show data across all birds (**a**) and for each bird individually (**b**). Overall, birds performed more poorly at all one-apart tasks compared to two- or three-apart tasks, perhaps suggesting that colours that are one step apart neared the physiological limit of colour discrimination. However, the pass rate for the 5|6 comparison (which crossed the category boundary; green box, grey-shaded area) was significantly higher than for comparisons that did not cross the category boundary (white-shaded areas; blue boxes) (paired *t*-test, $t_{24} = 4.09$, $P < 0.01$). Numbers in parentheses inside boxes are number of birds that participated in each type of comparison. Box plots depict the median (horizontal line inside the box), 25th and 75th percentiles (box), 25th and 75th percentiles $\pm 1.5\times$ interquartile range (whiskers) and outliers (circles). The horizontal grey line represents the expected pass frequency if birds flip discs by chance. Numbers in square brackets are Michelson contrast for a given colour pair.



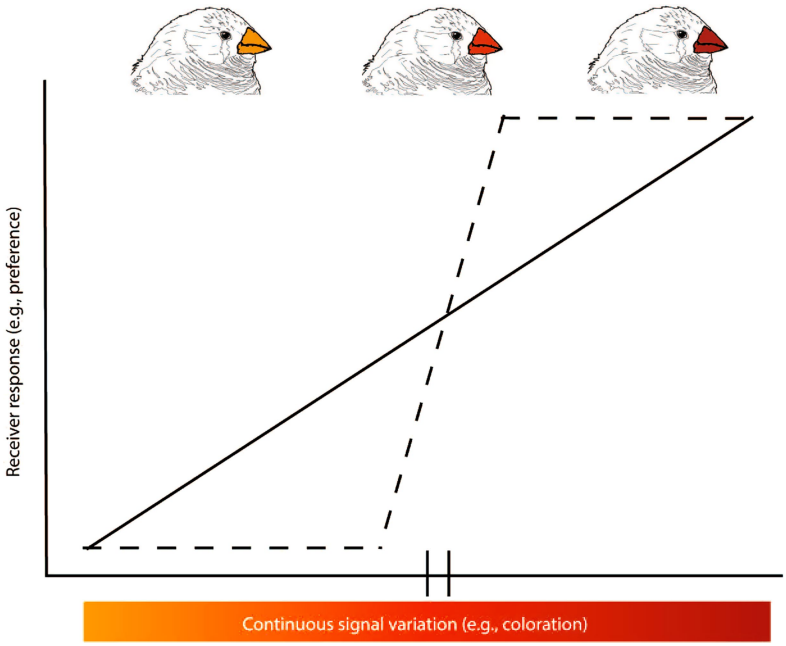
Extended Data Fig. 4 | Results from three-apart discrimination trials. **a**, **b**, Results show data across all birds (**a**) and for each bird individually (**b**). Pass rate was significantly higher for the three comparisons that crossed the boundary (3|6, 4|7 and 5|8; green boxes, grey-shaded area) than for those that did not (1|4 and 2|5; blue boxes, white-shaded area) (paired t-test, $t_{25} = 6.07$, $P < 0.001$). Numbers in parentheses inside boxes are number of birds that participated in each type of comparison. Box plots depict the median (horizontal line inside the box), 25th and 75th percentiles (box), 25th and 75th percentiles $\pm 1.5\times$ interquartile range (whiskers) and outliers (circles). The horizontal grey line represents the expected pass rate if birds flip discs by chance. Numbers in square brackets are Michelson contrast values for a given colour pair.



Extended Data Fig. 5 | Results from greyscale (that is, hue information removed) one-apart discrimination experiments. These greyscale experiments did not replicate the categories that we observed when hue information was included, indicating that categories are not structured based on brightness alone. Box plots depict the median (horizontal line inside the box), 25th and 75th percentiles (box), 25th and 75th percentiles ±1.5× interquartile range (whiskers) and outliers (circles). Numbers in parentheses below each comparison are Michelson contrast values. Linear mixed models showed that, in our greyscale experiments (that is, without chromaticity information), Michelson contrast between greyscale pairs significantly predicted pass rate. This finding is consistent with the possibility that category formation may be the result of both chromaticity and brightness information (see Table 1). Sample size was 18 birds for all comparisons.



Extended Data Fig. 6 | Wavelength discrimination of avian photoreceptors. The wavelength discrimination function of the pigeon *C. livia* (black dashed line) plotted against the spectral sensitivity peaks of the zebra finch (green, red, blue lines). Original data for the electroretinography-derived wavelength discrimination function are from Riggs et al.[31] and have been inverted so that higher numbers indicate greater discrimination.



Extended Data Fig. 7 | Schematic illustrating differences between continuous and categorical perception. Under continuous perception (solid line), receivers perceive and respond in a continuous fashion to signal variation, meaning that any change in a signalling trait is associated with a concomitant change in receiver response. Under categorical perception (dashed line), such as described here for female zebra finches, receivers show enhanced discrimination of variants across a boundary (hash marks on *x* axis) compared to equally spaced variants within a category. Zebra finch line drawing by N. Silina licensed under a Attribution 4.0 International (CC BY 4.0) licence (http://supercoloring.com/pages/zebra-finch).

Extended Data Table 1 | Photon catch values for the short-, medium- and long-wave cones of the zebra finch

Colour	S	M	L	Double
1 (5R 4/4)	7.73E+06	3.07E+07	9.89E+07	9.89E+07
2 (5R 4/6)	6.99E+06	3.01E+07	1.15E+08	1.08E+08
3 (7.5R 4/6)	6.08E+06	2.83E+07	1.15E+08	1.05E+08
4 (7.5R 4/8)	5.72E+06	2.83E+07	1.38E+08	1.16E+08
5 (10R 4/8)	4.50E+06	2.81E+07	1.38E+08	1.16E+08
6 (10R 5/10)	5.67E+06	4.27E+07	2.20E+08	1.83E+08
7 (10R 6/12)	7.42E+06	6.31E+07	3.35E+08	2.73E+08
8 (2.5YR 6/12)	6.26E+06	6.90E+07	3.42E+08	2.85E+08
Grey 1 (Carbon/Behr)	-	-	-	1.03E+08
Grey 2 (Deep Onyx/Glidden)	-	-	-	1.10E+08
Grey 3 (Black/Behr)	-	-	-	1.06E+08
Grey 4 (Night Club/Behr)	-	-	-	1.15E+08
Grey 5 (Night Club/Behr)	-	-	-	1.15E+08
Grey 6 (Flagstone Gray/Glidden)	-	-	-	1.82E+08
Grey 7 (Flipper/Behr)	-	-	-	2.77E+08
Grey 8 (Downtown Gray/Behr)	-	-	-	2.83E+08

Values were calculated using the reflectance spectrum of each Munsell colour or shade of grey, spectral sensitivity of the zebra finch single and double cones, and standard Illuminant A. Photon catches for Munsell colours are the average of four measurements and have not been relativized. For Munsell colours, the hue group and saturation/chroma are given. For greys, the name and brand of the paint swatch used are given.

Extended Data Table 2 | Chromatic distances between selected Munsell colours under different spectral sensitivities and lighting conditions

Condition	1v2	2v3	3v4	4v5	5v6	6v7
(1) Zebra Finch, Illuminant A	4.2	2.7	4.3	4.7	4.2	2.8
(2) Zebra finch, Daylight	4.3	2.7	4.1	4.8	4.1	2.7
(3) Zebra finch, experimental room lighting (halogen)	4.3	2.7	4.3	4.7	4.2	2.8
(4) Zebra finch, Von Kries adaptation, Illuminant A	4.5	2.3	3.9	4.0	3.5	2.2
(5) Starling, Illuminant A	3.8	1.8	3.5	3.5	3.5	2.1
(6) Average UVS, Illuminant A	3.8	1.8	3.5	3.5	3.5	2.1
(7) Average VS, Illuminant A	4.3	2.1	3.5	3.1	3.7	2.3

Chromatic distances were calculated using (1) zebra finch *Taeniopygia guttata* UVS cone-type retina and standard Illuminant A; (2) zebra finch spectral sensitivity and standard daylight Illuminant D65; (3) zebra finch spectral sensitivity, ambient light spectrum from experimental trials; (4) zebra finch spectral sensitivity, standard Illuminant A, accounting for Von Kries adaptation using a neutral grey background; (5) starling *S. vulgaris* UVS cone-type retina and standard Illuminant A; (6) the average avian UVS cone-type retina with standard Illuminant A; and (7) the average avian VS cone-type retina with standard Illuminant A.

Extended Data Table 3 | Results of a single linear mixed model containing all labelling and discrimination data

Parameter	Estimate	$\chi 2$ (1)	p value
Intercept	0.062		
One Colour Distance	0.020	147	<0.0001
Across 5v6 Boundary	0.21	39.1	<0.0001

As in the model presented in Table 1, this is a linear mixed effects model that includes bird ID as a random effect and the binary indicator (5|6) as random slopes.

LETTER

https://doi.org/10.1038/s41586-018-0399-1

Ecosystem warming extends vegetation activity but heightens vulnerability to cold temperatures

Andrew D. Richardson[1,2,3]*, Koen Hufkens[1], Thomas Milliman[4], Donald M. Aubrecht[1], Morgan E. Furze[1], Bijan Seyednasrollah[1,2,3], Misha B. Krassovski[5], John M. Latimer[5], W. Robert Nettles[5], Ryan R. Heiderman[5], Jeffrey M. Warren[5] & Paul J. Hanson[5]

Shifts in vegetation phenology are a key example of the biological effects of climate change[1–3]. However, there is substantial uncertainty about whether these temperature-driven trends will continue, or whether other factors—for example, photoperiod—will become more important as warming exceeds the bounds of historical variability[4,5]. Here we use phenological transition dates derived from digital repeat photography[6] to show that experimental whole-ecosystem warming treatments[7] of up to +9 °C linearly correlate with a delayed autumn green-down and advanced spring green-up of the dominant woody species in a boreal *Picea–Sphagnum* bog. Results were confirmed by direct observation of both vegetative and reproductive phenology of these and other bog plant species, and by multiple years of observations. There was little evidence that the observed responses were constrained by photoperiod. Our results indicate a likely extension of the period of vegetation activity by 1–2 weeks under a 'CO_2 stabilization' climate scenario (+2.6 ± 0.7 °C), and 3–6 weeks under a 'high-CO_2 emission' scenario (+5.9 ± 1.1 °C), by the end of the twenty-first century. We also observed severe tissue mortality in the warmest enclosures after a severe spring frost event. Failure to cue to photoperiod resulted in precocious green-up and a premature loss of frost hardiness[8], which suggests that vulnerability to spring frost damage will increase in a warmer world[9,10]. Vegetation strategies that have evolved to balance tradeoffs associated with phenological temperature tracking may be optimal under historical climates, but these strategies may not be optimized for future climate regimes. These in situ experimental results are of particular importance because boreal forests have both a circumpolar distribution and a key role in the global carbon cycle[11].

In temperate and boreal regions, rising temperatures are advancing spring onset (for example, budburst and flowering) and delaying autumn senescence (for example, leaf coloration and leaf fall)[12,13]. Whether these trends will be maintained is an open question[4]. Warm and cold temperatures, photoperiod and insolation, and precipitation and water availability have all been shown to influence plant phenology[2,5,14,15]. However, the future response of phenology to rising temperatures still remains largely unknown because of the high degree of uncertainty associated with interactions among these drivers[12]. Importantly, it has previously been proposed that photoperiod may constrain the phenological response to rising air temperatures[4,5,16]. Although there is evidence for this in some species[8,15], the generality of these results—and whether there are robust patterns across functional types—has yet to be demonstrated[5].

Analyses of observational datasets to disentangle the effects of these drivers are challenged by the lack of variability in natural systems, the inherent correlation among drivers and the realism of space-for-time assumptions[12]. Experimental approaches are thus required. However, there are sizable challenges associated with conducting realistic environmental manipulations, particularly for ecosystems with tall vegetation. Because of financial, logistical and technological hurdles, experimental warming treatments have not previously been applied to forest stands, and have only rarely been applied to single mature trees[17]. Although experiments with seedlings and branch cuttings are relatively common[18,19], artefacts associated with these approaches may limit their broader applicability[20,21].

We have been studying the effect of experimental whole-ecosystem warming treatments on vegetation phenology at the 'Spruce and Peatland Responses Under Changing Environments' (SPRUCE) facility, a long-term, multi-factor manipulative experiment situated in a boreal peatland forest in the Upper Midwest of the United States[7]. To our knowledge, this experiment is unique in that the five levels of warming (from 0 to +9 °C, see Methods, Extended Data Fig. 1, Supplementary Note 1, Extended Data Table 1) are being applied to intact communities of native plants, including woody shrubs and mature trees. The dominant plant species at SPRUCE represent key genera that are found across the vast boreal forest (taiga), which covers much of the land surface of the Northern Hemisphere from 45° to 70° N. Knowledge of the environmental controls on the phenology of these species is poor and does not at present provide a strong basis for making predictions about the capacity for phenological tracking of a warmer climate. Results from SPRUCE will therefore inform our understanding of the effects of climate change on processes related to biogeochemical cycling and biosphere–atmosphere feedbacks for this globally extensive biome.

Our focus here is on the effect of the experimental ecosystem warming treatments on spring and autumn phenology in this forested peat bog. Specifically, we tested three competing hypotheses: first, that temperature is the dominant control on phenological events (hereafter referred to as H1). This hypothesis predicts that the observed phenological transition date is directly related to the degree of warming (Fig. 1 a). Second, that photoperiod is the dominant control on phenological events (hereafter referred to as H2). This hypothesis predicts that the observed phenological transition date is constant regardless of the degree of warming (Fig. 1 b). Third, that photoperiod constrains the phenological response to temperature (hereafter referred to as H3). This hypothesis predicts that the observed response to temperature is flat beyond a threshold temperature, t^* (Fig. 1 c).

We tracked phenological responses to the experimental treatments in two ways. Since August 2015 we have monitored the vegetation within each enclosure using digital repeat photography[6] (Fig. 1 d, e), and since April 2016 we have made weekly ground observations of vegetative and reproductive phenology on a variety of plant species.

For our analysis of camera imagery, we distinguished between three distinct vegetation types: an evergreen conifer, *Picea mariana* (black spruce); a deciduous conifer, *Larix laricina* (eastern tamarack or larch); and a mixed, ground-level shrub community dominated by *Rhododendron groenlandicum* (Labrador tea) and *Chamaedaphne calyculata* (leatherleaf). For each vegetation type, green-down—as determined by G_{CC}, a colour index derived from the digital images—in

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA. [3]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA. [4]Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA. [5]Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA. *e-mail: Andrew.Richardson@nau.edu

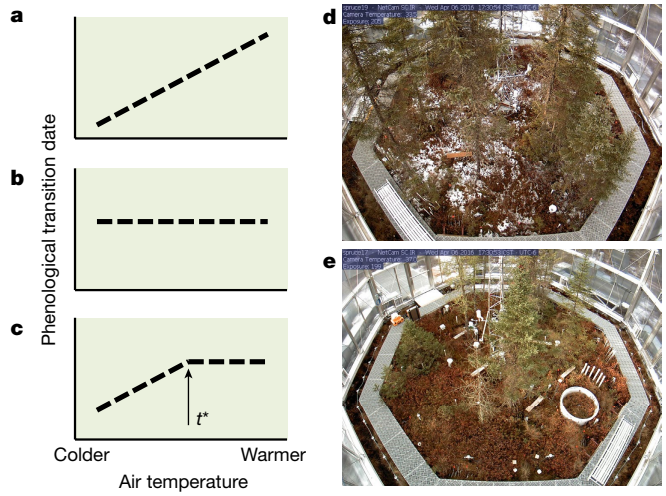


Fig. 1 | Testing competing hypotheses for phenological responses to warming using data from a whole-ecosystem warming experiment. **a**–**c**, Conceptual model of relationship between temperature and vegetation phenology, illustrating three competing hypotheses. **a**, Temperature is the dominant control (H1). **b**, Photoperiod is the dominant control (H2). **c**, Photoperiod limits the temperature response above the temperature threshold, t^* (H3). **d**, **e**, Sample digital camera imagery showing the inside of plot 19 (unheated control enclosure) (**d**) and plot 17 (+9.0 °C warming treatment enclosure) (**e**) on 6 April 2016. At the time the photographs were taken, the air temperature was 5 °C in plot 19 (note the last snow of the season), compared to 14 °C in plot 17.

autumn 2015 was delayed with increasing warming (Fig. 2a–c). The response to warming was significantly stronger (interaction effect between temperature and species, $P < 0.001$) for the mixed shrub community (about 5 days delay per 1 °C warming) than for either of the tree species (1–2 days delay per 1 °C warming), but was in all cases highly linear. Our results unequivocally support H1; that is, that temperature is the dominant control on the timing of autumn phenology. The fact that the temperature sensitivities were in all cases significantly different from zero allows us to reject H2. In no case did our breakpoint analysis (see Methods) identify a t^* value that substantially improved model fit (Extended Data Table 2), allowing us to reject H3. The above results are for autumn 2015, and comparable results were observed in autumn 2016 and 2017 (Supplementary Note 2).

Similarly, green-up in spring 2016 was advanced with increasing warming (Fig. 2d–f). The response to warming (1–2 days advancement per 1 °C warming) was not significantly different among vegetation types (interaction effect between temperature and species, $P = 0.34$). As in autumn, the fact that the temperature sensitivities were significantly different from zero allows us to reject H2. Breakpoint model analysis allowed us to reject H3, as in no case was a t^* value identified that would improve model fit (Extended Data Table 2). In spring, as in autumn, H1 is best supported by the experimental results. Results in spring 2017 were generally consistent with those for spring 2016 (Supplementary Note 2).

The above results clearly indicate a continued extension of the period of vegetation activity in response to future warming. By combining downscaled climate projections (Extended Data Fig. 2) from CMIP5[22] with the phenological temperature sensitivities estimated from Fig. 2 (Supplementary Note 3), we predict that the physiologically active season of the two conifer species may be extended by about a week under a 'CO_2 stabilization' climate scenario (representative concentration pathway (RCP)4.5, +2.9 ± 0.7 °C), and up to three weeks under a 'high CO_2 emission' scenario (RCP8.5, +5.9 ± 1.1 °C) by the year 2100 (Extended Data Table 3). Active season extension for the shrub layer is projected to be roughly twice as large as that of the conifers. These results are judged to be entirely plausible, given that future warming is not projected to exceed the levels of experimental warming at SPRUCE and that we are thus not extrapolating into unsampled climate space.

Previous work has shown that the seasonality of G_{CC} is a robust proxy for the seasonality of vegetation photosynthesis in both conifer forests and wetland ecosystems[23,24], and thus earlier plant green-up and delayed green-down at SPRUCE are almost certainly associated with a longer photosynthetically active period, and probably associated

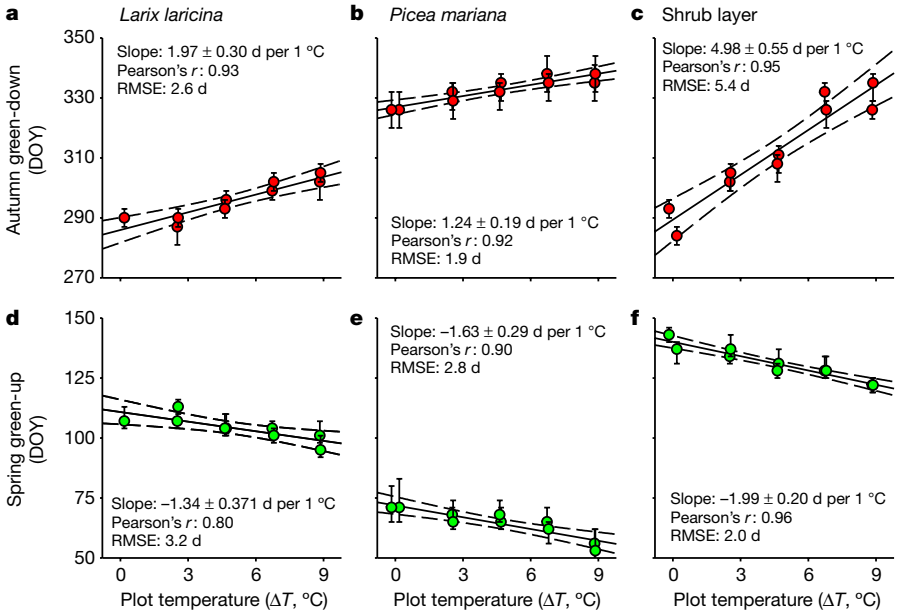


Fig. 2 | Effect of whole-ecosystem warming treatments on dates of autumn green-down and spring-green up, as derived from digital camera imagery. **a**–**f**, Response of autumn green-down (**a**–**c**, 2015) and spring green-up (**d**–**f**, 2016) phenology to experimental warming treatments for *L. laricina*, *P. mariana* and a mixed shrub layer community dominated by *R. groenlandicum* and *C. calyculata*, based on observations across $n = 10$ experimental enclosures ($n = 9$ for *Larix*, as in one enclosure this species was not within the camera field of view). Green-down and green-up are proxies for autumn senescence and spring onset, respectively. Error bars indicate 95% confidence interval around estimated phenological transition dates. Additional results are presented in Supplementary Note 2 and Extended Data Table 2. DOY, day of year; RMSE, root mean squared error.

Table 1 | Effect of SPRUCE warming treatments on spring and autumn phenological events (phenophases)

2016					2017			
Phenophase	*n*	*T* sensitivity	Breakpoints	*t**	*n*	*T* sensitivity	Breakpoints	*t**
Leaves growing	5	−3.19 ± 0.69	3	6.34 ± 1.08	7	−3.19 ± 1.25	5	5.46 ± 1.36
Shoots elongating	4	−3.53 ± 0.92	2	6.14 ± 1.59	5	−3.69 ± 1.04	4	7.12 ± 1.66
Flowering (cones open)	6	−2.51 ± 0.89	5	6.09 ± 1.89	7	−2.91 ± 1.33	5	6.74 ± 2.34
Flowers terminated					6	−1.45 ± 1.17	1	4.65
Fruiting	1	−2.56			6	−2.09 ± 1.38	3	6.06 ± 1.52
Autumn buds					3	−0.59 ± 1.03		
Autumn coloration (senescence)					6	2.70 ± 1.45	2	4.73 ± 2.88

Statistics (mean ± 1 s.d.) are calculated across all observed species. Sample size (*n*) indicates the number of species observed. '*T* sensitivity' is the temperature sensitivity, in days change per 1 °C warming, as estimated from the linear temperature model based on regression of transition date (*y*) on warming treatment (*x*). 'Breakpoints' indicates the number of species for which a *t** threshold was identified using the breakpoint temperature model (response is linear up to a temperature threshold *t**, and flat thereafter; see Fig. 1c and Methods). Species-level results are reported in Supplementary Note 5 and Extended Data Tables 4, 5.

with enhanced annual photosynthetic uptake (though not necessarily increased vegetation growth). This result is consistent with the analysis of long-term data from FLUXNET sites (http://fluxnet.fluxdata.org/, Supplementary Note 4, Extended Data Fig. 3), as well as previous experimental[19] and observational[25] studies. However, this does not necessarily indicate an increase in net carbon uptake or carbon sequestration under future warming, because the long-term carbon balance of this peatland forest ecosystem is probably dependent on the stability of the underlying peat deposits[26].

Camera-based results are generally consistent with direct observation of spring (2016 and 2017) and autumn (2017 only) phenological transitions for plant species spanning a range of leaf habits and growth forms (Table 1; see also Supplementary Note 5, Extended Data Tables 4, 5). Spring phenophases advanced by just over three days per 1 °C warming, providing strong support for H1. Autumn phenophases related to leaf coloration or senescence were delayed by almost three days per 1 °C warming, again providing support for H1. Relatively little variation was observed in dates of autumn bud set for *Chamaedaphne* and *Picea*, providing support for H2 for this particular phenophase of these species. Although *t** breakpoints that improved model fit were commonly identified, we note that in most cases the small-sample-corrected Akaike's information criterion (ΔAIC_C; see Methods) was greater than zero, which means that the simpler, linear temperature model was better supported by the data. Furthermore, the identified breakpoint temperatures were generally very high—below 4.5 °C in only a few instances—indicating that future warming would have to greatly exceed RCP4.5 projections before photoperiod constraints begin to limit phenological shifts. The ground observations therefore robustly support H1 over H2 or H3, and are consistent with the future extension of the active season at both ends.

There is abundant evidence in the literature that photoperiod has a role in triggering phenological events[27,28]. In many species, there has been a local adaptation of phenology to both photoperiod and temperature cues[5,15]. In some species and environments, photoperiod sets a hard limit on the phenological response to rising temperatures[4,15]. But, with warming of up to +9 °C above current levels, we found little evidence for this in most of the species and phenophases that we studied. Thus, photoperiod requirements are still being met even during the shortened winter simulated by the warmest enclosures. In the few cases in which there was evidence of a photoperiod effect, it was generally only a factor at temperatures well above current temperatures, again indicating that substantial future warming would be required for photoperiod to become limiting. These findings are consistent with a recent analysis showing that for high-latitude species, spring leaf-out was generally not sensitive to photoperiod[8].

The purported role of photoperiod as a phenological constraint is to prevent plants from responding to temperature signals at the 'wrong' time of the year[4]. However, if photoperiod is not a strong constraint on spring phenological development, then a counterintuitive prediction is that continued warming coupled with increasing frequency of climate extremes may increase the likelihood of spring frost damage[9,10]. At SPRUCE, atypical weather in March (unusually warm) and April (extreme cold) 2016 showed that in addition to triggering visually apparent phenological shifts, the warming treatments also advanced tissue de-hardening and thereby heightened the potential for spring frost damage (Supplementary Note 6, Extended Data Fig. 4). Following a spring frost event in which ambient temperatures dropped to −15 °C, we observed extensive foliar damage in the +9.0 °C enclosures (in which temperatures dropped to about −4 °C) and moderate damage in the +6.75 °C enclosures. Minimal damage occurred in the enclosures that received less warming and thus experienced colder minimum temperatures. This suggests that the transition from frost-hardy to frost-vulnerable is cued by warm temperatures[9], and is not constrained by photoperiod. Without photoperiod as a safety check on the de-hardening process, frost damage may be more severe and/or more frequent under future climate conditions. Woody plants generally have sufficient nonstructural carbon reserves to recover from occasional frost damage[10], but repeated damage could impair the competitive ability of susceptible species[9,29] (Extended Data Table 6).

Results from the first two-and-a-half years of the SPRUCE experiment, conducted in a winter-dormant ecosystem, show decisively that warming treatments directly influence vegetation phenology at both the start and end of the annual period of vegetation activity. These phenological shifts will almost certainly influence photosynthesis and transpiration[3,16], as well as feedbacks to the climate system through effects on the surface energy budget[12]. Future extension of the active season in most cases appears unlikely to be strongly constrained by photoperiod in this boreal ecosystem. Potentially inopportune responses to environmental signals may occur as the climate moves beyond the range of historical variability, as demonstrated by the spring frost damage in the warmest enclosures. Thus, temperature-tracking strategies evolved to guide phenological responses to historical year-to-year variation in weather may be increasingly mismatched to future conditions[5].

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0399-1.

Received: 15 August 2017; Accepted: 28 June 2018;
Published online 8 August 2018.

1. Settele, J. et al. in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Field, C. B. et al.) 271–359 (Cambridge Univ. Press, Cambridge, 2014).
2. Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A. & Schwartz, M. D. Shifting plant phenology in response to global change. *Trends Ecol. Evol.* **22**, 357–365 (2007).
3. Morisette, J. T. et al. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. *Front. Ecol. Environ.* **7**, 253–260 (2009).

4. Körner, C. & Basler, D. Phenology under global warming. *Science* **327**, 1461–1462 (2010).
5. Way, D. A. & Montgomery, R. A. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant Cell Environ.* **38**, 1725–1736 (2015).
6. Sonnentag, O. et al. Digital repeat photography for phenological research in forest ecosystems. *Agric. For. Meteorol.* **152**, 159–177 (2012).
7. Hanson, P. J. et al. Attaining whole-ecosystem warming using air and deep-soil heating methods with an elevated CO_2 atmosphere. *Biogeosciences* **14**, 861–883 (2017).
8. Zohner, C. M., Benito, B. M., Svenning, J.-C. & Renner, S. S. Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nat. Clim. Change* **6**, 1120–1123 (2016).
9. Augspurger, C. K. Reconstructing patterns of temperature, phenology, and frost damage over 124 years: spring damage risk is increasing. *Ecology* **94**, 41–50 (2013).
10. Gu, L. et al. The 2007 Eastern US spring freeze: increased cold damage in a warming world. *Bioscience* **58**, 253–262 (2008).
11. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**, 1444–1449 (2008).
12. Richardson, A. D. et al. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric. For. Meteorol.* **169**, 156–173 (2013).
13. Gill, A. L. et al. Changes in autumn senescence in northern hemisphere deciduous trees: a meta-analysis of autumn phenology studies. *Ann. Bot.* **116**, 875–888 (2015).
14. Laube, J. et al. Chilling outweighs photoperiod in preventing precocious spring development. *Glob. Change Biol.* **20**, 170–182 (2014).
15. Basler, D. & Körner, C. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agric. For. Meteorol.* **165**, 73–81 (2012).
16. Migliavacca, M. et al. On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. *Biogeosciences* **9**, 2063–2083 (2012).
17. Slaney, M., Wallin, G., Medhurst, J. & Linder, S. Impact of elevated carbon dioxide concentration and temperature on bud burst and shoot growth of boreal Norway spruce. *Tree Physiol.* **27**, 301–312 (2007).
18. Gunderson, C. A. et al. Forest phenology and a warmer climate – growing season extension in relation to climatic provenance. *Glob. Change Biol.* **18**, 2008–2025 (2012).
19. Stinziano, J. R., Hüner, N. P. A. & Way, D. A. Warming delays autumn declines in photosynthetic capacity in a boreal conifer, Norway spruce (*Picea abies*). *Tree Physiol.* **35**, 1303–1313 (2015).
20. Vitasse, Y. & Basler, D. Is the use of cuttings a good proxy to explore phenological responses of temperate forests in warming and photoperiod experiments? *Tree Physiol.* **34**, 174–183 (2014).
21. Morin, X., Roy, J., Sonié, L. & Chuine, I. Changes in leaf phenology of three European oak species in response to experimental climate change. *New Phytol.* **186**, 900–910 (2010).
22. Brekke, L., Thrasher, B., Maurer, E. & Pruitt, T. *Downscaled CMIP3 and CMIP5 Climate Projections* (US Department of the Interior, Bureau of Reclamation, Technical Services Center, 2013).
23. Peichl, M., Sonnentag, O. & Nilsson, M. B. Bringing color into the picture: using digital repeat photography to investigate phenology controls of the carbon dioxide exchange in a boreal mire. *Ecosystems* **18**, 115–131 (2015).
24. Bowling, D. R. et al. Limitations to winter and spring photosynthesis of a Rocky Mountain subalpine forest. *Agric. For. Meteorol.* **252**, 241–255 (2018).
25. Richardson, A. D. et al. Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests. *Tree Physiol.* **29**, 321–331 (2009).
26. Wilson, R. M. et al. Stability of peatland carbon to rising temperatures. *Nat. Commun.* **7**, 13723 (2016).
27. Singh, R. K., Svystun, T., AlDahmash, B., Jönsson, A. M. & Bhalerao, R. P. Photoperiod- and temperature-mediated control of phenology in trees – a molecular perspective. *New Phytol.* **213**, 511–524 (2017).
28. Zohner, C. M. & Renner, S. S. Perception of photoperiod in individual buds of mature trees regulates leaf-out. *New Phytol.* **208**, 1023–1030 (2015).
29. Hufkens, K. et al. Ecological impacts of a widespread frost event following early spring leaf-out. *Glob. Change Biol.* **18**, 2365–2377 (2012).

Acknowledgements This material is based upon work supported by the US Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for DOE under contract DE-AC05-00OR22725. Support for PhenoCam has come from the National Science Foundation (EF-1065029, EF-1702697). D. Hollinger, M. Carbone and C. Iverson provided feedback on a draft manuscript. E. Ward assisted with litter collection. For CMIP, we acknowledge the World Climate Research Programme's Working Group on Coupled Modelling. We thank the climate modelling groups (listed in Supplementary Note 3) for making their model output available. DOE's Program for Climate Model Diagnosis and Intercomparison additionally provides coordinating support and led development of software infrastructure for CMIP in partnership with the Global Organization for Earth System Science Portals.

Reviewer information *Nature* thanks M. Tjoelker and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.D.R. designed the study with input from P.J.H. A.D.R., K.H., D.M.A., T.M., M.E.F., B.S. and M.B.K. contributed PhenoCam imagery and derived data. J.M.L., W.R.N., J.M.W. and R.R.H. contributed phenological observations. J.M.W. contributed data on frost damage. M.B.K., W.R.N. and P.J.H. maintained site infrastructure including warming treatments and meteorological observations. A.D.R. assembled datasets and conducted the analysis. A.D.R. drafted the manuscript. All authors commented on and approved the final manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0399-1.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0399-1.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.D.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Statistical methods were not used to predetermine sample size for the regression design. The warming treatments were randomized among 10 plots with similar vegetation and uniform peat depths. Investigators were not blinded to allocation during experiments and outcome assessment.

Study site and experimental design. The SPRUCE experiment is located within the S1 peat bog at the Marcell Experimental Forest (47° 30.171′ N, 93° 28.970′ W)[30], approximately 40 km north of Grand Rapids in north-central Minnesota. The historical climate at the site is sub-humid continental: mean annual temperature is 4 °C, mean annual precipitation is 750 mm, and extreme temperatures range from −38 °C to +30 °C. Because this ecosystem is located at the southern edge of the boreal zone, it is considered particularly vulnerable to climate change.

The S1 bog is an ombotrophic peatland with a perched water table. Trees are approximately 5–8 m in height. Canopy vegetation is dominated by the tree species *P. mariana* (Mill.) B.S.P. (black spruce), with additional contributions from *L. laricina* (Du Roi) K. Koch (eastern tamarack or larch). *P. mariana* and *L. laricina* both have a vast geographic range across North America, from Alaska east to Quebec and Labrador, and south to the Great Lakes and New England. A number of closely related *Picea* and *Larix* species are distributed across the boreal zone of northern Europe, Scandinavia and much of Russia and Siberia, indicating the relevance of results of this experiment to our understanding of boreal ecosystem processes globally.

The SPRUCE understory is dominated by the evergreen shrubs *R. groenlandicum* (Oeder) Kron and Judd (Labrador tea) and *C. calyculata* (L.) Moench. (leatherleaf), and is underlain by a bryophyte layer dominated by *Sphagnum* spp. moss. Other common plant species include the evergreen shrub *Kalmia polifolia* Wangenh. (bog laurel), the deciduous shrub *Vaccinium angustifolium* Aiton 1789 not Benth. 1840 (lowbush blueberry), the sedge *Eriophorum* spp. (cottongrass), and the perennial herb *Maianthemum trifolium* (L.) Sloboda (false Solomon's seal).

At SPRUCE, experimental temperature (+0 °C 'unheated control' to +9.0 °C, in 2.25 °C increments for both air and deep soil) and CO_2 (ambient and elevated, approximately 400 and 900 p.p.m., respectively) treatments are being applied through the use of large (approximately 12-m wide, 8-m high) open-topped octagonal enclosures[7]. Overall, five temperature treatments are paired with two CO_2 treatments, yielding a total of ten enclosures (additionally, there are two 'ambient environment' plots without constructed enclosures). Each enclosure is hydrologically isolated from the rest of the bog by a sheet pile corral which has been driven 3–4 m through the peat into the underlying ancient lake sediments. Outflow pipes allow for lateral drainage from each enclosure. Within each enclosure, warming of the deep soil began in June 2014, while aboveground warming was initiated in August 2015 and at this time the phenological observations were commenced in each individual plot (note that pre-treatment observations were made in a common area, outside of the enclosures, beginning in 2010). CO_2 treatments were switched on in June 2016.

For context, the warmest enclosures (+9.0 °C) simulate current climate conditions of Wichita, Kansas (mean annual temperature 13 °C, mean annual precipitation 850 mm), located approximately 1,100 km (10° of latitude) to the south. The SPRUCE experiment, with treatments that will exceed the historical range of climatic variability (Extended Data Fig. 1), is intentionally planned to push the system past projected warming levels to approach or include tipping points for any number of ecosystem response variables. The regression-based experimental design facilitates the estimation of temperature response functions, which may be nonlinear[7].

The enclosure design, and detailed performance metrics for the above- and belowground warming, along with a discussion of potential artefacts, are more fully described and assessed in a previous publication[7]. Observed temperature differentials were consistent with the nominal warming treatments for target enclosures. Warming was homogeneous within individual enclosures, and was sustained over time (see Supplementary Note 1, Extended Data Table 1).

Phenological observations. We are using two methods to track the phenological responses of vegetation to warming and elevated CO_2 in each enclosure. First, beginning in August 2015, we installed digital cameras[31], or phenocams[32], in each enclosure to track seasonal variation in vegetation 'greenness', a proxy for vegetation phenology and associated physiological activity[6,33–35]. Second, beginning in April 2016, human observers have been directly tracking phenological events of both woody and herbaceous species.

PhenoCam imagery. Digital cameras (NetCam model SD130BN, StarDot Technologies) were configured and installed following standard protocols of the PhenoCam network[36]. Cameras record sequential visible-light (red, green, blue; RGB) and visible + infrared images[37] every 30 min from 4:00 to 22:00, every day of the year. Minimally compressed JPEG images, accompanied by a metadata file containing the current status of all camera settings and diagnostics, are uploaded via file transfer protocol to the PhenoCam server for archiving and processing; a local copy is also maintained on a server running at SPRUCE. The filename of every image identifies the enclosure in which the picture was recorded, as well as a date and time stamp in local standard time.

The aluminium structural members of each enclosure provided convenient and consistent mounting points for the cameras. All cameras were mounted, at a height of 6 m, in the middle of the third horizontal structural member on the south wall of each enclosure. Cameras were enclosed in lightweight, compact weatherproof enclosures (model ENC-OUTD3, StarDot Technologies). Network connectivity and DC power were delivered to each camera using a single Ethernet cable and standard power-over-Ethernet technology. To reduce the likelihood of lightning damage, an Ethernet surge protector (ProtectNet model PNET1GB, APC by Schneider Electric) was installed on the camera end of each Ethernet cable, and grounded to the mounting point.

All imagery is posted in near-real time to the PhenoCam project web page (http://phenocam.sr.unh.edu/), where it is publicly available. Images are processed nightly, using standard PhenoCam routines[6,36]. In brief, this consists of several steps. First, we defined three separate regions of interest (ROIs) for each camera field of view, demarcating (1) *Picea* trees; (2) *Larix* trees; and (3) the mixed shrub layer. The ROI definitions are converted to binary masks, so that image analysis can be completed separately for each vegetation type. Next, images were read in sequentially, and for each vegetation type the mean pixel value for each of the three colour channels (red, green and blue; for the purposes of the present analysis we used only the visible-wavelength imagery) was calculated across the corresponding ROI, yielding a digital number (DN) triplet (R_{DN}, G_{DN}, B_{DN}). Then for each ROI in each image, we calculated the green chromatic coordinate G_{CC}, which has previously been shown to be a reliable metric for characterizing the seasonal trajectory of vegetation colour and activity[6,31,38]:

$$G_{CC} = \frac{G_{DN}}{R_{DN} + G_{DN} + B_{DN}}$$

Basic quality control included eliminating images that were recorded when the sun was less than 5° above the horizon, images that were too dark or images that were too bright. Additionally, because snow might obscure the vegetation of interest, for each day from late August 2015 through the end of December 2017, we visually inspected the mid-day image from each camera. We flagged images in which there was (1) snow on the ground; or (2) snow on trees. We excluded from further processing all days on which the camera's view of the vegetation of interest was potentially contaminated by snow. For the shrub layer, this meant eliminating images from days with snow on the ground; for *Picea* and *Larix*, this meant eliminating images from days with snow on trees. The frequency of snow decreased with increasing plot temperature, from over 100 days per year with snow on the ground in the unheated enclosures (from late October to early May), to less than 30 days per year in the +9.0 °C enclosures (from late November to early February). The longest period of continuous snow cover was almost three months in the unheated enclosures, compared with only two weeks in the +9.0 °C enclosures.

Next, we determined 3-day G_{CC} values using the 90th quantile method[6]. We then used a spline-based method to sequentially remove outliers in three iterative steps. Finally, we re-fit the spline, and used the summertime maxima and dormant-season minima to define the seasonal G_{CC} amplitude, from which we were then able to identify dates at which 10%, 25% and 50% of the seasonal amplitude were reached in autumn (senescent or green-down phase) and spring (onset or green-up phase). Uncertainties on these dates were then derived based on the uncertainty around the smoothing spline. Our analysis here focuses on the 25% amplitude threshold dates.

Ground observations. Ground observations of spring phenology were made at approximately weekly intervals by W.R.N. and J.M.L. in 2016, and by R.R.H. in 2017. The protocol used by W.R.N. and R.R.H. involved recording, on a pre-printed form for each of the 10 enclosures and the two ambient environment plots, whether or not ('yes' or 'no') specific vegetative and reproductive phenophases were observed each week. Observations were conducted on a selection of woody species (the trees *Picea* and *Larix*; the evergreen shrubs leatherleaf, bog laurel, Labrador tea and lowbush blueberry), as well as a sedge (cottongrass) and a perennial herb (false Solomon's seal). We transcribed the data by taking as the observed date the first survey date on which an event was definitively observed (that is, 'no' through week 4, followed by 'yes' in week 5: the event occurred in week 5). Not all phenophases were observed for all species, and in some difficult-to-observe cases, the data were deemed not reliable because of some inconsistencies in the recorded data (for example, blank cells rather than 'no', or 'no' followed by 'yes' followed by 'no' again) or poor representation of the species in question in some of the plots (for example, bog laurel and lowbush blueberry are sparsely distributed). All transcribed data of questionable reliability were excluded from the analysis.

J.M.L.'s protocol involved recording the first date at which *Larix* leaf buds were observed to be just beginning to break (data recorded for all ten enclosures,

plus the two ambient environment plots), and the first date on which flowers of leatherleaf, bog laurel and Labrador tea were observed in each enclosure (data recorded in only half of the treated enclosures, plus one or both of the ambient environment plots). Although data recorded by J.M.L. are not as complete as those recorded by W.R.N., they are included to demonstrate the robustness of the observed patterns.

On-site meteorological data. Air temperature and relative humidity were measured (model HMP-155, Vaisala) at four points above the peat surface within each enclosure (0.5, 1, 2 and 4 m), and 30-min mean values recorded. We used the measured air temperature at 2 m in our analyses. SPRUCE environmental data[39] are available through the Vista Data Vision portal (http://sprucedata.ornl.gov).

Historical perspective and future climate projections. To put the weather during winter and spring of 2016 in historical context (122 year record), we used data from the National Climatic Data Center of the NOAA. Specifically, we used summary data from the State of the Climate report (https://www.ncdc.noaa.gov/sotc/national/), and three-month divisional temperature rankings (https://www.ncdc.noaa.gov/temp-and-precip/climatological-rankings/). The SPRUCE site falls within Minnesota's climate division 2.

To place our results in the context of projected warming trends over the coming century, we used downscaled (1/8°) climate projections from a selection of ten models (see Supplementary Note 2) contributing to the CMIP5 multimodel ensemble dataset[22,40]. We used output for two RCP scenarios: RCP4.5 (CO_2 stabilization) and RCP8.5 (high CO_2 emission)[41,42]. To quantify future trends, we calculated the projected decadal mean air temperature change relative to the 2006–2015 mean for each model.

Statistical analysis. To characterize the relationship between air temperature and phenological timing (H1 and H2), we used ordinary linear regression, with the observed phenological date as the dependent variable, y_i, and the measured air temperature differential for each plot (see Supplementary Note 1) as the independent variable, x_i. The regression slope β thus gives the temperature sensitivity in days per 1 °C warming for the 'linear temperature model'. To account for potential effects of elevated CO_2 on phenology, we also analysed data (where appropriate) using a 'linear temperature and CO_2 model', which included temperature, CO_2 (elevated and ambient) and a temperature × CO_2 interaction effect. All tests were two-sided, at a significance level of 0.05.

For breakpoint analysis (H3), we fit a three-parameter (α, β and t^*) 'breakpoint temperature model', which was specified as:

$$y_i = \alpha + \beta x_i + \epsilon_i \text{ for } x_i < t^*$$

and

$$y_i = \alpha + \beta t^* + \epsilon_i \text{ for } x_i \geq t^*$$

in which x_i and y_i are as for the ordinary linear regression, ϵ_i is the regression residual and t^* is the temperature breakpoint, as illustrated in Fig. 1. We constrained t^* to fall in the range of 2–9 °C. An edge-hitting value of $t^* = 9$ °C was obtained when the linear model fit the data every bit as well as the breakpoint model.

We used AIC[43] to identify whether the linear model or the breakpoint model was best supported by the available data. AIC is typically calculated as:

$$\text{AIC} = n\log\sigma^2 + 2p$$

in which n is the number of observations, p is the number of fit parameters plus one, and σ^2 is the residual sum of squares divided by n. When n is small relative to p, the small-sample-corrected criterion, AIC_C, is preferred[43]:

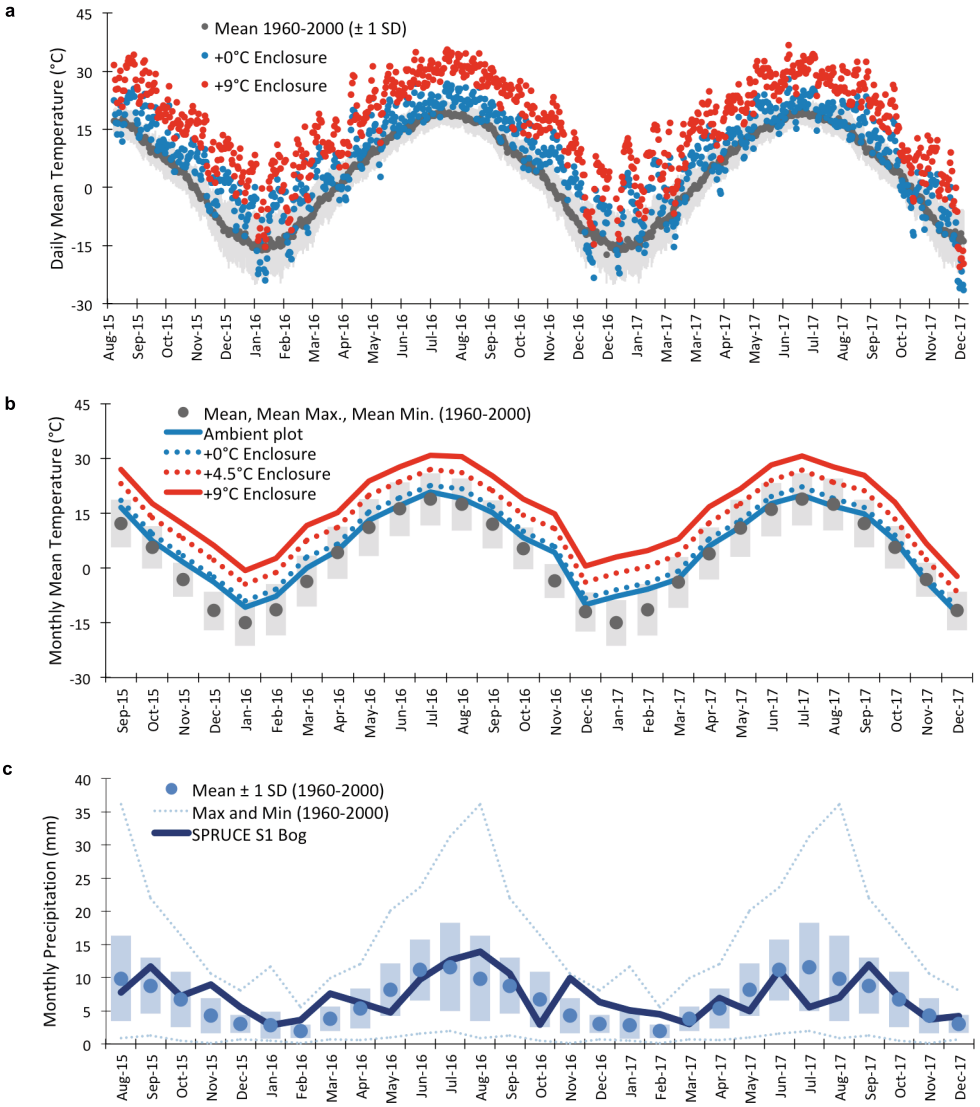
$$\text{AIC}_C = \text{AIC} + \frac{2p(p+1)}{n-p-1}$$

AIC effectively balances improving explanatory power (lower σ^2) against increasing complexity (larger p), and thus AIC selects against over-parameterized models. The model with the lowest AIC is considered the best model given the data, and the absolute difference in AIC_C scores between two models can be used to evaluate the weight of evidence in support of the better model. If the difference (ΔAIC) is small or zero then the two models are equally good. But, if $\Delta\text{AIC} \approx 2.0$, then the model with the lower AIC_C is almost three times more likely to be the best[43].

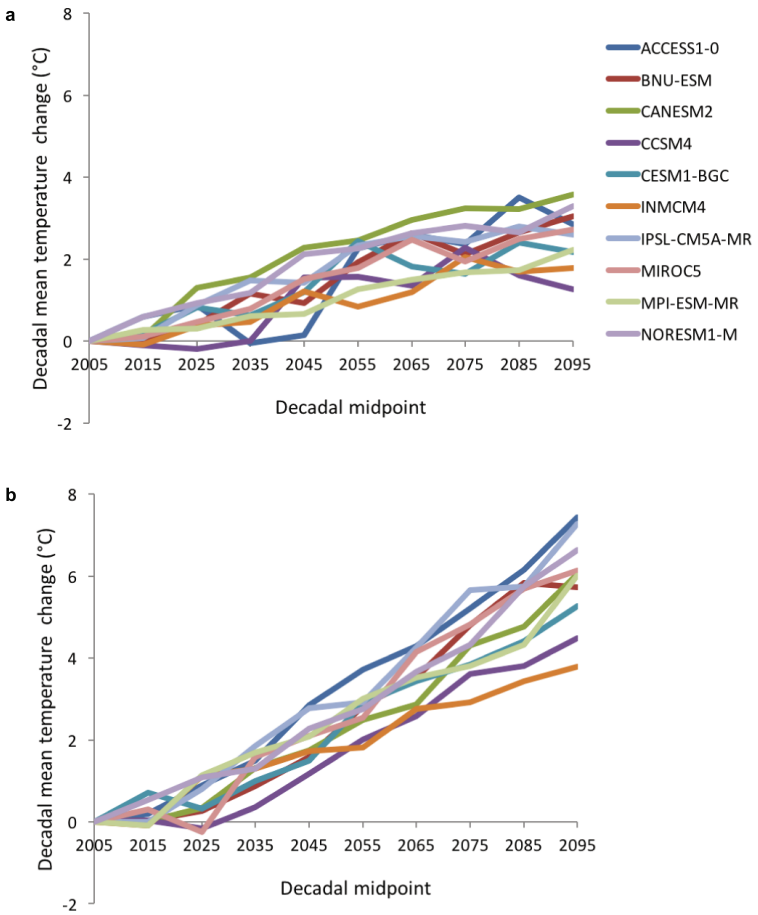
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. PhenoCam imagery is publicly available through the project web page (http://phenocam.sr.unh.edu), and the phenological datasets used in this study are available through the SPRUCE data portal[44,45].

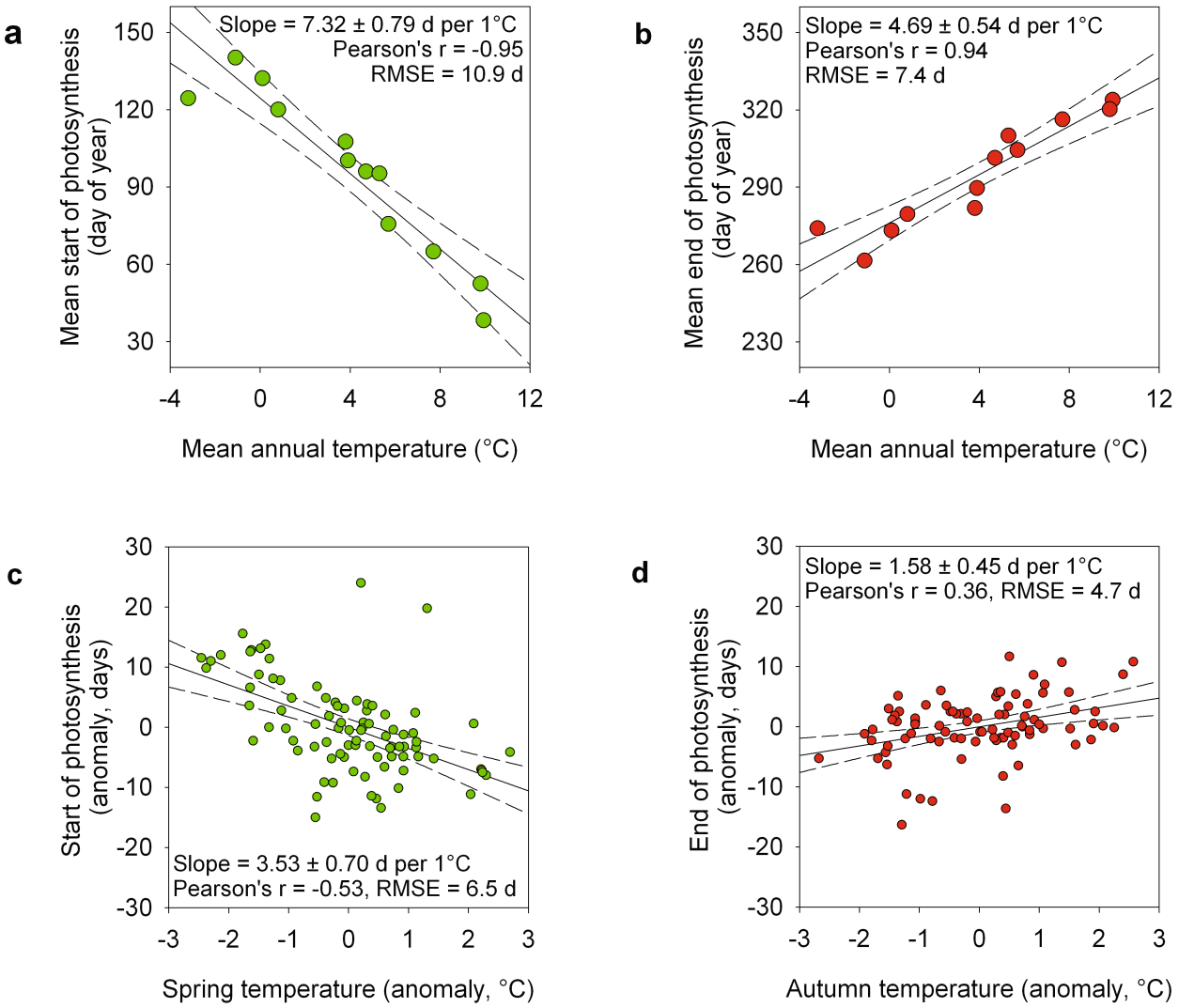
30. Kolka, R., Sebestyen, S., Verry, E. S. & Brooks, K. *Peatland Biogeochemistry and Watershed Hydrology at the Marcell Experimental Forest* (CRC, Boca Raton, 2011).
31. Richardson, A. D. et al. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia* **152**, 323–334 (2007).
32. Brown, T. B. et al. Using phenocams to monitor our changing Earth: toward a global phenocam network. *Front. Ecol. Environ.* **14**, 84–93 (2016).
33. Keenan, T. F. et al. Tracking forest phenology and seasonal physiology using digital repeat photography: a critical assessment. *Ecol. Appl.* **24**, 1478–1489 (2014).
34. Richardson, A. D., Klosterman, S. & Toomey, M. in *Phenology: An Integrative Environmental Science* (ed. Schwartz, M. D.) 413–430 (Springer, Dordrecht, 2013).
35. Toomey, M. et al. Greenness indices from digital cameras predict the timing and seasonal dynamics of canopy-scale photosynthesis. *Ecol. Appl.* **25**, 99–115 (2015).
36. Richardson, A. D. et al. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Sci. Data* **5**, 180028 (2018).
37. Petach, A. R., Toomey, M., Aubrecht, D. M. & Richardson, A. D. Monitoring vegetation phenology using an infrared-enabled security camera. *Agric. For. Meteorol.* **195–196**, 143–151 (2014).
38. Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Jenkins, J. P. & Ollinger, S. V. Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecol. Appl.* **19**, 1417–1428 (2009).
39. Hanson, P., Riggs, J., Nettles, W., Krassovski, M. & Hook, L. *SPRUCE Whole Ecosystems Warming (WEW) Environmental Data Beginning August 2015* https://doi.org/10.3334/CDIAC/spruce.032 (2016).
40. Maurer, E. P., Brekke, L., Pruitt, T. & Duffy, P. B. Fine-resolution climate projections enhance regional climate change impact studies. *Eos* **88**, 504 (2007).
41. Riahi, K. et al. RCP 8.5—a scenario of comparatively high greenhouse gas emissions. *Clim. Change* **109**, 33–57 (2011).
42. Thomson, A. M. et al. RCP4.5: a pathway for stabilization of radiative forcing by 2100. *Clim. Change* **109**, 77–94 (2011).
43. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: a Practical Information–Theoretic Approach* (Springer, New York, 2002).
44. Richardson, A. D. et al. *SPRUCE Ground Observations of Phenology in Experimental Plots 2016–2017* https://doi.org/10.3334/CDIAC/spruce.044 (2018).
45. Richardson, A. D. et al. *SPRUCE Vegetation Phenology in Experimental Plots from Phenocam Imagery 2015–2017* https://doi.org/10.3334/CDIAC/spruce.045 (2018).



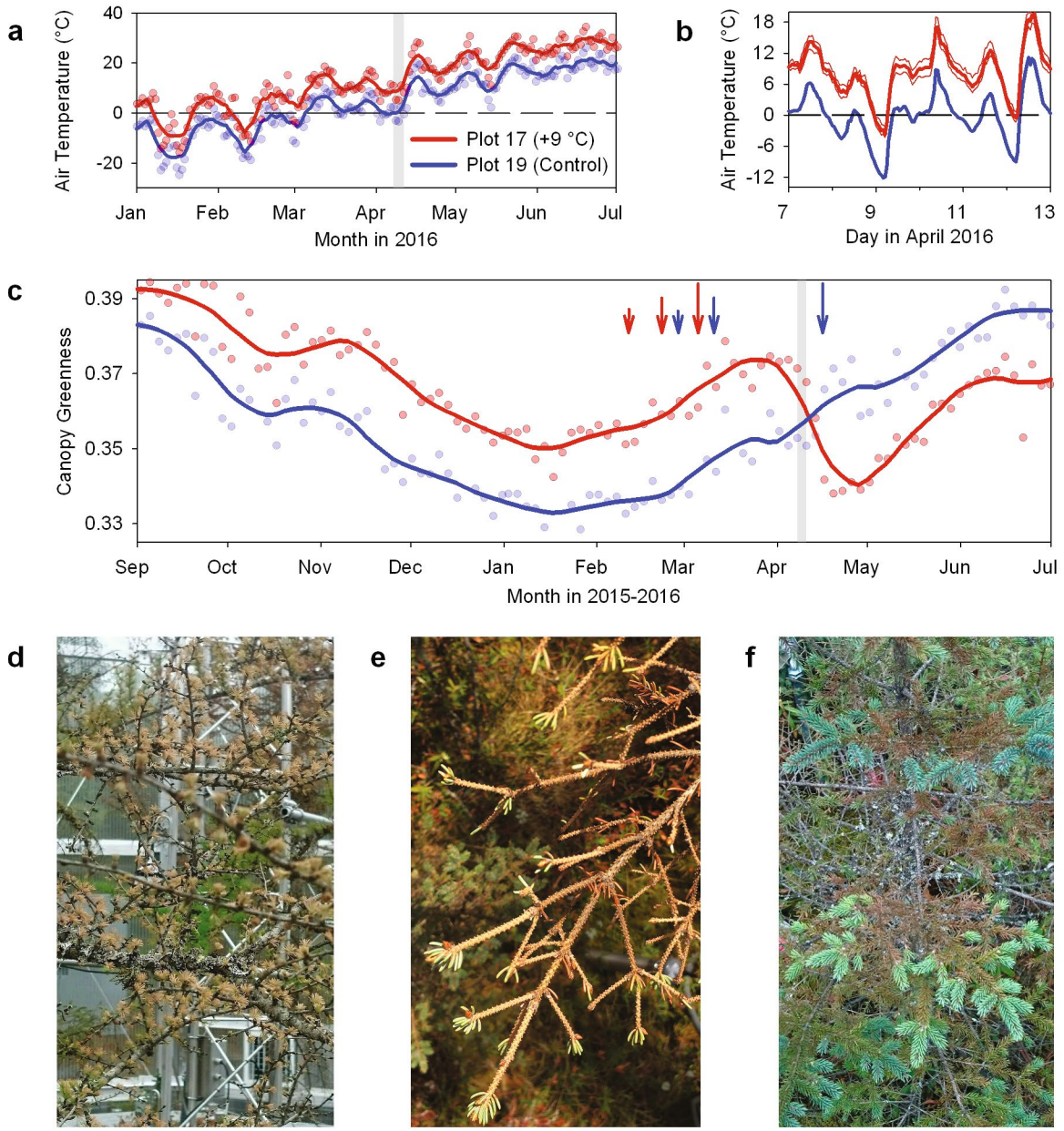
Extended Data Fig. 1 | Air temperature and precipitation in the SPRUCE S1 bog (August 2015 to December 2017) relative to long-term (1960–2000) means and variability. a, Long-term daily mean temperature (°C, ± 1 s.d. indicated by shading), compared with daily mean temperature (calculated from 30-min means, based on $n = 2$ sensors mounted at 2-m height in each enclosure) in a +0 °C enclosure (unheated control) and a +9.0 °C enclosure. **b**, Long-term monthly mean temperature (mean daily maximum and mean daily minimum indicated by shaded bars), compared with monthly mean temperature (calculated from daily means, as in **a**) in different experimental treatments. **c**, Long-term monthly mean precipitation (mm, ± 1 s.d. indicated by shading, with maxima and minima indicated by dotted lines), compared with measured monthly precipitation ($n = 1$ rain gauge) in the S1 bog.



Extended Data Fig. 2 | Decadal mean temperature change (relative to 2006–2015 mean) projections from ten CMIP5 earth system models for the SPRUCE site. a, Stabilization climate scenario (RCP4.5). **b**, High emission climate scenario (RCP8.5).



Extended Data Fig. 3 | Relationships between air temperature and the start and end of the photosynthetic uptake period, as derived from FLUXNET data for evergreen conifer-dominated sites. a–d, Across-site patterns in spring (**a**) and autumn (**b**) in relation to mean annual temperature ($n = 12$ sites), and within-sites patterns in spring (**c**) and autumn (**d**) in relation to seasonal temperature anomalies ($n = 86$ site-years).



Extended Data Fig. 4 | Unusually warm weather in late winter, followed by extreme cold in early April, resulted in severe frost damage in the warmest enclosures at SPRUCE in 2016. a, Time series of daily mean air temperature, comparing plot 17 (+9.0 °C warming) and plot 19 (unheated enclosure), during the winter and spring of 2016. By the time the frost event occurred (grey shading), the daily mean temperature in plot 17 had been above freezing for over a month, but had repeatedly dropped below freezing in plot 19. **b**, Time series of 30-min air temperature—again comparing plot 17 and plot 19—leading up to and immediately following the frost event, which occurred on the morning of 9 April and again on 12 April. The thin red lines indicate the variability (maximum and minimum) across $n = 5$ temperature sensors in plot 17. **c**, Time series of daily G_{CC}, the green chromatic coordinate, for *Picea* trees in plot 17 and plot 19. Arrows denote spring green-up dates (progressively larger arrows corresponding to 10%, 25% and 50% of seasonal amplitude) estimated from G_{CC}. The pronounced decline in G_{CC} in plot 17 following the frost event (grey shading) is readily apparent. Trees in plot 19 retained sufficient frost hardiness that they were undamaged, despite experiencing much colder temperatures. **d**, Brown frost-damaged *Larix* foliage in plot 17. **e**, *Picea* branches in plot 17, showing loss of most foliage from previous years, with green foliage from the 2015 flush retained only at branch tips. **f**, *Picea* branches with frost-damaged foliage from previous years, but healthy green foliage from the 2016 flush.

Extended Data Table 1 | Mean daily air temperature and temperature differentials associated with whole-ecosystem warming

Plot number	Plot 21	Plot 07	Plot 19	Plot 06	Plot 11	Plot 20	Plot 13	Plot 04	Plot 08	Plot 16	Plot 10	Plot 17
Nominal Treatment	Ambient	Ambient	Unheated	Unheated	+2.25	+2.25	+4.5	+4.5	+6.75	+6.75	+9.0	+9.0
Mean daily mean temperature (°C)												
Aug-Dec 2015	7.0	6.8	8.2	8.6	10.8	10.9	12.9	13.1	15.2	15.3	17.1	17.2
Jan-Jun 2016	2.9	2.9	4.5	4.9	7.2	7.4	9.5	9.4	11.4	11.4	13.4	13.5
Jul-Dec 2016	9.6	9.7	11.1	11.4	13.8	13.8	15.9	15.9	17.9	18.0	20.1	20.2
Jan-Jun 2017	2.8	3.0	4.5	4.8	7.3	7.2	9.3	9.4	11.5	11.5	13.6	13.6
Jul-Dec 2017	9.8	9.8	11.3	11.5	13.9	13.9	15.9	16.0	18.0	18.2	20.3	20.3
Mean temperature differential (ΔT, °C)												
Aug-Dec 2015	-1.4	-1.6	-0.2	0.2	2.4	2.5	4.5	4.7	6.8	6.9	8.7	8.8
Jan-Jun 2016	-1.8	-1.8	-0.2	0.2	2.5	2.7	4.7	4.7	6.6	6.6	8.7	8.8
Jul-Dec 2016	-1.6	-1.5	-0.2	0.2	2.5	2.5	4.6	4.6	6.6	6.8	8.9	8.9
Jan-Jun 2017	-1.8	-1.7	-0.2	0.2	2.6	2.5	4.7	4.7	6.9	6.9	8.9	8.9
Jul-Dec 2017	-1.6	-1.6	-0.1	0.1	2.6	2.5	4.6	4.6	6.7	6.8	9.0	9.0
Mean ΔT ± 1 SD	-1.6 ± 0.2	-1.6 ± 0.1	-0.2 ± 0.0	0.2 ± 0.0	2.5 ± 0.1	2.6 ± 0.1	4.6 ± 0.1	4.7 ± 0.1	6.7 ± 0.1	6.8 ± 0.1	8.8 ± 0.1	8.9 ± 0.1

Daily means are calculated on the basis of the mean half-hour data for two temperature sensors mounted at 2-m height. Temperature differentials (ΔT) are calculated relative to the mean of the two unheated enclosures (plots 19 and 6). Plots are arranged in order of increasing ΔT; overall mean $\pm$ 1 standard deviation (SD). ΔT is calculated across $n = 5$ multi-month means.

Extended Data Table 2 | Effect of SPRUCE warming treatments on spring green-up and autumn green-down

			Linear Temperature Model				Breakpoint Temperature Model			
Species	Phenophase	Mean DOY	Pearson *r*	RMSE	Slope ± 1 SE	T effect	RMSE	Slope ± 1 SE	*t** ± SE	ΔAICc
Larix laricina	Autumn 2015	296 ± 6	0.93	2.6	1.97 ± 0.30	P < 0.001	-	-	-	-
	Spring 2016	104 ± 5	-0.80	3.2	-1.34 ± 0.37	P < 0.01	-	-	-	-
	Autumn 2016	298 ± 10	0.64	8.1	2.12 ± 0.95	P = 0.06	-	-	-	-
	Spring 2017	101 ± 4	-0.81	2.6	-1.11 ± 0.30	P < 0.01	2.5	-3.15 ± 1.27	3.50 ± 0.90	+5.1
	Autumn 2017	293 ± 11	0.66	9.0	2.48 ± 1.07	P = 0.05	-	-	-	-
Picea marinana	Autumn 2015	333 ± 4	0.92	1.9	1.24 ± 0.19	P < 0.001	1.6	1.61 ± 0.35	6.43 ± 1.08	+1.6
	Spring 2016	64 ± 6	-0.90	2.8	-1.63 ± 0.29	P < 0.001	-	-	-	-
	Autumn 2016	325 ± 9	0.86	5.1	2.47 ± 0.52	P < 0.01	5.3	2.82 ± 0.74	7.48 ± 1.81	+5.4
	Spring 2017	88 ± 12	0.19	12.2	0.69 ± 1.25	P = 0.60	-	-	-	-
	Autumn 2017	315 ± 7	0.91	3.2	2.00 ± 0.33	P < 0.001	3.1	2.42 ± 0.43	6.90 ± 1.17	+3.5
Shrub layer	Autumn 2015	312 ± 17	0.95	5.4	4.98 ± 0.55	P < 0.0001	5.3	5.64 ± 0.74	7.55 ± 0.91	+4.0
	Spring 2016	131 ± 7	-0.96	2.0	-1.99 ± 0.20	P < 0.0001	-	-	-	-
	Autumn 2016	313 ± 19	0.92	7.5	5.25 ± 0.77	P < 0.001	7.4	6.08 ± 1.04	7.31 ± 1.16	+4.2
	Spring 2017	128 ± 7	-0.67	5.8	-1.48 ± 0.59	P = 0.04	-	-	-	-
	Autumn 2017	305 ± 13	0.94	4.8	3.66 ± 0.49	P < 0.0001	3.4	5.49 ± 0.73	5.59 ± 0.58	-2.0

Results (derived from PhenoCam imagery) are shown from the start of the whole-ecosystem warming experiment (autumn 2015), on the basis of observations across $n = 10$ experimental enclosures ($n = 9$ for *Larix*, as in one enclosure this species was not within the camera field of view). Mean transition dates are reported ± 1 s.d. Statistics for the linear temperature model are based on regression of transition date (y) on warming treatment (x), and the model slope is the phenological temperature sensitivity in days per 1 °C warming. The '*T* effect' column reports the *P* value for the null hypothesis of no temperature effect. Statistics for the breakpoint temperature model are based on a model in which the response to warming treatment is assumed to be linear up to a temperature threshold t^*, and flat thereafter (see Methods for additional details). No statistics are reported for cases in which a t^* could not be identified or where the addition of t^* did not improve model fit. ΔAIC_C is the difference in AIC (corrected for small sample sizes) between the linear temperature model and the breakpoint temperature model, with a positive value indicating that the linear temperature model is better supported by the data and a negative value indicating that the beakpoint temperature model is better supported by the data. RMSE, root mean squared error. SE, standard error. Results not shown for the linear temperature and CO_2 model as the CO_2 effect and $CO_2 \times T$ interaction effect were generally not significant (see Supplementary Note 2 for additional information).

Extended Data Table 3 | Projected future extension of the period of vegetation activity

Temperature sensitivity (days per 1°C warming)		*Larix laricina*	*Picea mariana*	Shrub layer
Onset of green-up		-1.3 ± 0.4	-1.6 ± 0.3	-2.0 ± 0.2
Completion of green-down		2.0 ± 0.3	1.2 ± 0.2	5.0 ± 0.6
Temperature sensitivity of total active season length		3.3	2.9	7.0
Projected active season extension (days)				
RCP 4.5	Extension by 2055	6 ± 2	5 ± 2	13 ± 4
	Extension by 2095	9 ± 2	7 ± 2	18 ± 5
RCP 8.5	Extension by 2055	9 ± 2	8 ± 1	19 ± 3
	Extension by 2095	20 ± 4	17 ± 3	41 ± 8

The model is based on linear extrapolation of experimental results, using CMIP5 climate projections. Temperature sensitivities are derived from Fig. 2; total projected active season extension is the product of the temperature sensitivity of total active season length multiplied by the mean projected temperature increase (decadal means, relative to 2006–2015). Uncertainties in active season extension represent the uncertainty in the climate projections (s.d. across ten models) rather than the uncertainty in the temperature sensitivities.

Extended Data Table 4 | Effect of SPRUCE warming treatments on observed vegetative and reproductive phenological transitions (2016)

			Linear temperature model				Breakpoint temperature model			
Species	Phenophase	Mean DOY	Pearson *r*	RMSE	Slope ± SE	T effect	RMSE	Slope ± SE	*t** ± SE	ΔAICc
Picea mariana	Buds breaking	140 ± 10	0.97	2.8	-2.72 ± 0.23	P < 0.0001	2.9	-2.87 ± 0.31	7.73 ± 0.98	+4.5
(Black spruce, evergreen tree)	Shoots elongating	142 ± 9	0.93	3.6	-2.35 ± 0.29	P < 0.0001	-	-	-	-
Larix laricina	Buds breaking	95 ± 15	0.98	3.6	-4.06 ± 0.29	P < 0.0001	3.3	-4.42 ± 0.35	7.30 ± 0.71	+1.7
(Eastern tamarack, deciduous tree)	Leaves growing	110 ± 10	0.93	3.8	-2.45 ± 0.30	P < 0.0001	3.8	-2.66 ± 0.40	7.31 ± 1.34	+3.8
	Shoots elongating	149 ± 16	0.94	6.0	-4.13 ± 0.49	P < 0.0001	-	-	-	-
C. calyculata	Flowering	110 ± 10	0.97	2.7	-2.63 ± 0.22	P < 0.0001	2.4	-2.93 ± 0.25	6.97 ± 0.75	+0.9
(Leatherleaf, evergreen shrub)	Leaves growing	126 ± 18	0.91	7.6	-4.33 ± 0.62	P < 0.0001	6.1	-6.01 ± 0.90	5.18 ± 0.85	-1.7
	Shoots elongating	128 ± 17	0.93	6.7	-4.37 ± 0.55	P < 0.0001	6.8	-4.78 ± 0.72	7.26 ± 1.32	+3.6
	Flowering (2nd obs.)	108 ± 10	0.83	6.1	-2.00 ± 0.61	P = 0.02	-	-	-	-
Kalmia polifolia	Flowering	129 ± 14	0.75	10.0	-2.83 ± 0.84	P < 0.01	9.4	-5.57 ± 2.18	3.26 ± 1.55	+2.5
(Bog laurel, evergreen shrub)	Flowering (2nd obs.)	125 ± 15	0.95	5.4	-3.18 ± 0.48	P < 0.01	-	-	-	-
R. groenlandicum	Buds breaking	121 ± 10	0.92	4.3	-2.53 ± 0.35	P < 0.0001	-	-	-	-
(Labrador tea, evergreen shrub)	Leaves growing	127 ± 13	0.89	6.2	-3.12 ± 0.51	P < 0.001	-	-	-	-
	Shoots elongating	129 ± 14	0.86	7.7	-3.26 ± 0.62	P < 0.001	6.9	-4.64 ± 1.02	5.01 ± 1.22	+1.0
	Flowering	130 ± 15	0.92	6.2	-3.65 ± 0.50	P < 0.0001	5.9	-5.03 ± 0.86	5.05 ± 0.95	+2.1
	Flowering (2nd obs.)	121 ± 16	1.00	0.7	-3.56 ± 0.06	P < 0.0001	0.6	-3.61 ± 0.06	8.36 ± 0.19	+9.7
V. angustifolium	Buds breaking	117 ± 13	0.98	2.8	-3.34 ± 0.25	P < 0.0001	2.9	-3.61 ± 0.42	7.69 ± 0.98	+5.1
(Lowbush blueberry, deciduous shrub)	Leaves growing	119 ± 13	0.87	6.8	-3.08 ± 0.55	P < 0.001	-	-	-	-
	Flowering	119 ± 11	0.96	3.3	-2.84 ± 0.27	P < 0.0001	3.4	-2.99 ± 0.36	7.78 ± 1.09	+4.1
Eriophorum spp.	Flowering	107 ± 5	0.74	3.6	-1.03 ± 0.29	P < 0.01	-	-	-	-
(Cottongrass, sedge)	Seeds formed	124 ± 10	0.96	2.9	-2.56 ± 0.24	P < 0.0001	-	-	-	-
M.trifolium	Leaves growing	137 ± 13	0.87	6.5	-2.95 ± 0.53	P < 0.001	6.5	-3.41 ± 0.78	6.52 ± 1.65	+3.4
(False Solomon's seal, perennial herb)	Flowering	148 ± 9	0.86	4.8	-2.06 ± 0.39	P < 0.001	5.0	-2.23 ± 0.53	7.37 ± 2.12	+4.4

Data are from 2016 growing season, based on observations across $n = 12$ plots. Species are ordered by functional type, and within each species, phenophases are ordered according to the mean (± 1 s.d.) day of year (DOY) on which the event occurred. Statistics for the linear temperature model are based on regression of transition date (y) on warming treatment (x), and the model slope is the phenological temperature sensitivity in days per 1 °C warming. The '*T* effect' column reports the *P* value for the null hypothesis of no temperature effect. Statistics for the breakpoint temperature model are based on a model in which the response to warming treatment is assumed to be linear up to a temperature threshold t^*, and flat thereafter (see Methods for additional details). No statistics are reported for cases in which a t^* could not be identified, or where the addition of t^* did not improve model fit. ΔAIC_C is the difference in AIC (corrected for small sample sizes) between the linear temperature model and breakpoint temperature model, with a positive value indicating that the linear temperature model is better supported by the data, and a negative value indicating that the breakpoint temperature model is better supported by the data. RMSE, root mean squared error. SE, standard error.

Extended Data Table 5 | Effect of SPRUCE warming treatments on observed vegetative and reproductive phenological transitions (2017)

			Linear Temperature Model				Breakpoint Temperature Model			
Species	Phenophase	Mean DOY	Pearson *r*	RMSE	Slope ± SE	T effect	RMSE	Slope ± SE	*t** ± SE	ΔAICc
C. calyculata	Flowering	107 ± 6	0.91	3.0	-1.46 ± 0.21	P < 0.0001	-	-	-	-
	Leaves growing	134 ± 25	0.90	11.0	-5.75 ± 0.88	P < 0.0001	8.3	-7.41 ± 1.22	6.26 ± 1.06	-3.6
	Shoots elongating	143 ± 20	0.95	7.0	-4.85 ± 0.52	P < 0.0001	7.0	-4.91 ± 0.72	8.76 ± 1.44	+4.7
	Fruiting	147 ± 5	0.71	3.0	-0.86 ± 0.27	P = 0.01	3.3	-1.45 ± 0.78	4.31 ± 2.53	+2.5
	Flowers terminated	159 ± 11	0.78	7.0	-2.17 ± 0.55	P < 0.01	-	-	-	-
	Fall buds	249 ± 0	0.00	0.0	0.00 ± 0.00	-	-	-	-	-
	Leaves colored	310 ± 11	0.88	5.0	2.48 ± 0.42	P < 0.001	3.5	5.79 ± 0.82	2.69 ± 0.49	-6.9
Eriophorum spp.	Leaves greening	105 ± 16	0.81	10.0	-3.38 ± 0.78	P < 0.01	8.4	-5.39 ± 1.23	4.77 ± 1.23	-0.8
	Flowering	106 ± 9	0.66	7.0	-1.50 ± 0.54	P = 0.02	-	-	-	-
	Fruiting	143 ± 24	0.66	19.0	-4.15 ± 1.48	P = 0.02	-	-	-	-
	Flowers terminated	187 ± 29	0.16	30.0	-1.21 ± 2.36	P = 0.62	-	-	-	-
Kalmia polifolia	Leaves greening	113 ± 29	0.82	17.0	-6.22 ± 1.35	P < 0.01	13.8	-12.43 ± 3.23	3.50 ± 1.05	-1.9
	Flowering	120 ± 14	0.95	5.0	-3.38 ± 0.35	P < 0.0001	3.6	-4.52 ± 0.53	5.67 ± 0.70	-2.0
	Leaves growing	135 ± 20	0.57	17.0	-2.96 ± 1.35	P = 0.05	-	-	-	-
	Fruiting	153 ± 11	0.77	7.0	-2.31 ± 0.63	P < 0.01	-	-	-	-
	Flowers terminated	158 ± 9	0.89	4.0	-1.99 ± 0.32	P < 0.0001	-	-	-	-
Larix laricina	Leaves growing	110 ± 11	0.84	6.0	-2.30 ± 0.47	P < 0.001	3.9	-4.93 ± 0.91	3.27 ± 0.71	-7.0
	Shoots elongating	147 ± 11	0.93	4.0	-2.72 ± 0.33	P < 0.0001	4.3	-2.95 ± 0.44	7.76 ± 1.38	+3.9
	Fall buds	260 ± 22	0.31	23.0	-1.78 ± 1.96	P = 0.39	-	-	-	-
	Leaves yellow	300 ± 12	0.94	4.0	2.85 ± 0.33	P < 0.0001	-	-	-	-
	Leaves dropped	318 ± 16	0.93	6.0	3.91 ± 0.49	P < 0.0001	-	-	-	-
M. trifolium	Leaves growing	135 ± 9	0.92	4.0	-2.12 ± 0.28	P < 0.0001	-	-	-	-
	Flowering	140 ± 8	0.90	4.0	-1.94 ± 0.30	P < 0.0001	3.3	-3.99 ± 0.77	3.10 ± 0.72	-0.4
	Fruiting	162 ± 8	0.12	9.0	-0.27 ± 0.74	P = 0.73	-	-	-	-
	Flowers terminated	202 ± 23	0.13	24.0	0.76 ± 1.85	P = 0.69	-	-	-	-
	Leaves senesced	294 ± 17	0.62	14.0	2.84 ± 1.21	P = 0.04	-	-	-	-
Picea mariana	Cones open	134 ± 24	0.82	14.0	-5.01 ± 1.11	P < 0.01	14.8	-5.36 ± 1.53	7.97 ± 2.66	+4.5
	Leaves growing	143 ± 14	0.93	5.0	-3.31 ± 0.41	P < 0.0001	4.9	-4.58 ± 0.72	5.34 ± 0.91	+1.8
	Shoots elongating	155 ± 11	0.91	5.0	-2.64 ± 0.39	P < 0.0001	3.6	-4.13 ± 0.52	4.66 ± 0.67	-4.9
	Fall buds	249 ± 0	0.00	0.0	0.00 ± 0.00	-	-	-	-	-
R. groenlandicum	Flowering	126 ± 15	0.98	3.0	-3.83 ± 0.27	P < 0.0001	3.6	-3.95 ± 0.37	8.47 ± 0.91	+4.3
	Leaves growing	134 ± 11	0.79	7.0	-2.34 ± 0.57	P < 0.01	7.0	-2.83 ± 0.73	6.85 ± 2.23	+2.6
	Shoots elongating	136 ± 14	0.98	3.0	-3.56 ± 0.25	P < 0.0001	2.6	-4.00 ± 0.27	7.30 ± 0.60	-1.7
	Fruiting	154 ± 9	0.96	3.0	-2.28 ± 0.21	P < 0.0001	2.4	-2.62 ± 0.24	7.02 ± 0.82	-0.1
	Flowers terminated	167 ± 7	0.87	4.0	-1.59 ± 0.28	P < 0.001	-	-	-	-
	Last flowers (2nd flush)	277 ± 44	0.74	32.0	9.33 ± 3.22	P = 0.02	-	-	-	-
	Leaves folding	307 ± 19	0.84	11.0	4.06 ± 0.84	P < 0.001	9.8	4.93 ± 1.44	6.77 ± 1.99	+1.2
V. angustifolium	Leaves growing	118 ± 15	0.90	7.0	-3.54 ± 0.55	P < 0.0001	5.7	-4.90 ± 0.83	5.60 ± 1.01	-1.7
	Flowering	120 ± 15	0.85	8.0	-3.23 ± 0.62	P < 0.001	8.4	-3.33 ± 0.86	8.47 ± 2.49	+4.7
	Shoots elongating	140 ± 25	0.75	17.0	-4.66 ± 1.36	P < 0.01	-	-	-	-
	Fruiting	145 ± 15	0.69	11.0	-2.65 ± 0.88	P = 0.01	11.6	-3.07 ± 1.20	6.86 ± 3.39	+4.3
	Flowers terminated	161 ± 15	0.67	11.0	-2.51 ± 0.89	P = 0.02	11.3	-3.94 ± 1.93	4.65 ± 2.49	+3.4
	Fruits ripened	181 ± 15	0.90	7.0	-3.57 ± 0.59	P < 0.001	4.6	-5.84 ± 1.08	4.18 ± 0.88	-2.9
	Leaves colored	303 ± 19	0.01	20.0	0.04 ± 1.54	P = 0.98	-	-	-	-

Data are from the 2017 growing season, based on observations across $n = 12$ plots. Species are ordered alphabetically, and within each species, phenophases are ordered according to the mean (± 1 s.d.) day of year (DOY) on which the event occurred. Statistics for the linear temperature model are based on regression of transition date (y) on warming treatment (x), and the model slope is the phenological temperature sensitivity in days per 1 °C warming. The '*T* effect' column reports the *P* value for the null hypothesis of no temperature effect. Statistics for the breakpoint temperature model are based on a model in which the response to warming treatment is assumed to be linear up to a temperature threshold t^*, and flat thereafter (see Methods for additional details). No statistics are reported for cases in which a t^* could not be identified, or where the addition of t^* did not improve model fit. ΔAIC_C is the difference in AIC (corrected for small sample sizes) between the linear temperature model and breakpoint temperature model, with a positive value indicating that the linear temperature model is better supported by the data, and a negative value indicating that the breakpoint temperature model is better supported by the data. RMSE, root mean squared error. SE, standard error. Results not shown for the linear temperature and CO_2 model as the CO_2 effect and $CO_2 \times T$ interaction effect were generally not significant (see Supplementary Note 5 for additional information).

Extended Data Table 6 | Impact of premature foliar senescence on nutrient content of *L. laricina* and *P. mariana* litter

		Larix laricina		*Picea mariana*	
		mean	range	mean	range
Spring, post-frost	N (%)	1.79	(0.71-2.85)	0.71	(0.43-0.83)
Premature senescent litter	C (%)	46.2	(46.1-46.3)	46.3	(44.7-50.0)
Fall	N (%)	0.35	(0.30-0.40)	0.42	(0.34-0.50)
Normal senescent litter	C (%)	51.1	(50.9-51.4)	51.8	(51.1-52.5)
Senescent litter ratio	N_{pre}:N_{norm}	5.11		1.69	
Premature:Normal	C_{pre}:C_{norm}	0.90		0.89	

Following the 9 April 2016 spring frost event, damaged foliage from trees that had lost frost hardiness began a period of senescence, culminating in heavy leaf fall during early May as air temperatures frequently exceeded 30 °C in the +9.0 °C plots (temperatures over 40 °C were observed in plot 10 and plot 17 on 5 and 6 May). Prematurely senescent litter was collected on 6 May from the ground underneath damaged trees in the two warmest treatments (+6.75 and +9.0 °C) (n = 3–7 trees). Normally senescent litter was collected on 4 November from ambient environment plots outside of the experimental treatments using litter baskets (n = 8 trees). Litter was analysed for carbon and nitrogen by combustion using 0.1-g samples of oven-dried and finely ground tissue on a TruSpec elemental analyser (LECO). Data are presented on a per cent dry matter basis.

LETTER

https://doi.org/10.1038/s41586-018-0360-3

Accumulation of 8,9-unsaturated sterols drives oligodendrocyte formation and remyelination

Zita Hubler[1,8], Dharmaraja Allimuthu[1,8], Ilya Bederman[2], Matthew S. Elitt[1], Mayur Madhavan[1], Kevin C. Allan[1], H. Elizabeth Shick[1], Eric Garrison[3], Molly T. Karl[3], Daniel C. Factor[1], Zachary S. Nevin[1], Joel L. Sax[1], Matthew A. Thompson[1], Yuriy Fedorov[4], Jing Jin[5], William K. Wilson[5], Martin Giera[6], Franz Bracher[7], Robert H. Miller[3], Paul J. Tesar[1] & Drew J. Adams[1]*

Regeneration of myelin is mediated by oligodendrocyte progenitor cells—an abundant stem cell population in the central nervous system (CNS) and the principal source of new myelinating oligodendrocytes. Loss of myelin-producing oligodendrocytes in the CNS underlies a number of neurological diseases, including multiple sclerosis and diverse genetic diseases[1–3]. High-throughput chemical screening approaches have been used to identify small molecules that stimulate the formation of oligodendrocytes from oligodendrocyte progenitor cells and functionally enhance remyelination in vivo[4–10]. Here we show that a wide range of these pro-myelinating small molecules function not through their canonical targets but by directly inhibiting CYP51, TM7SF2, or EBP, a narrow range of enzymes within the cholesterol biosynthesis pathway. Subsequent accumulation of the 8,9-unsaturated sterol substrates of these enzymes is a key mechanistic node that promotes oligodendrocyte formation, as 8,9-unsaturated sterols are effective when supplied to oligodendrocyte progenitor cells in purified form whereas analogous sterols that lack this structural feature have no effect. Collectively, our results define a unifying sterol-based mechanism of action for most known small-molecule enhancers of oligodendrocyte formation and highlight specific targets to propel the development of optimal remyelinating therapeutics.

Imidazole antifungal drugs are a structurally diverse class of small molecules that robustly stimulate the generation of new mouse and human oligodendrocytes and enhance remyelination in mouse models of disease[4]. Imidazole antifungals mediate their effects in yeast by inhibiting CYP51, an essential enzyme for sterol biosynthesis in both fungal and mammalian cells (for a detailed cholesterol biosynthesis diagram, see Extended Data Fig. 1). Across a panel of nine azole-containing molecules, the ability to inhibit CYP51 in vitro and in oligodendrocyte progenitor cells (OPCs) predicted enhanced formation of myelin basic protein-positive (MBP^+) oligodendrocytes from mouse epiblast stem cell-derived OPCs (Fig. 1a–d, Extended Data Fig. 2a–c). To measure inhibition of CYP51 in OPCs, we used gas chromatography and mass spectrometry (GC–MS) to quantify the increase in levels of lanosterol (the substrate of CYP51) and decrease in cholesterol[11–13] (Fig. 1b, Extended Data Fig. 2c–e). In cells treated with ketoconazole, the dose–response curve for accumulation of lanosterol closely resembled the dose–response curve for enhanced oligodendrocyte formation (Fig. 1c, Extended Data Fig. 2b, f, g). Notably, we confirmed all effects of small molecules on oligodendrocyte formation and sterol levels using a second, independently isolated batch of OPCs, and key results were also validated using mouse primary OPCs (Extended Data Fig. 2b–i; see Methods). In addition, the effects of azole molecules were confirmed using an orthogonal image quantification approach, a second oligodendrocyte marker, and liquid chromatography with mass spectrometry (LC–MS) to detect cellular sterols (Extended Data Fig. 2j–l).

We next used RNA interference and metabolite supplementation to independently confirm the role of CYP51 in oligodendrocyte formation. Cell-permeable small interfering RNA (siRNA) reagents depleted *CYP51* transcript levels in OPCs by 80%[14], led to substantial accumulation of lanosterol, and enhanced formation of MBP^+ oligodendrocytes (Fig. 1e, f, Extended Data Fig. 2m–o). In addition, we treated OPCs directly with purified lanosterol and observed enhanced formation of MBP^+ oligodendrocytes in a dose-responsive fashion (Fig. 1g, h, Extended Data Fig. 2p, q). These findings support the idea that CYP51 is the functional target of imidazole antifungals in OPCs and suggest that accumulation of sterol intermediates may play a direct role in enhancing oligodendrocyte formation.

As inhibition of CYP51 was sufficient to induce the formation of oligodendrocytes, we used a chemical genetics approach to test whether modulation of other steps in cholesterol biosynthesis had a similar effect (Fig. 2a, Extended Data Fig. 1). We used GC–MS-based sterol profiling in OPCs to confirm that a panel of eight small molecules selectively inhibited their known enzyme targets within the cholesterol biosynthesis pathway (Extended Data Fig. 3a–d; see Source Data for abundance of all quantified metabolites in all GC–MS-based sterol profiling experiments). Only molecules targeting CYP51 (ketoconazole), TM7SF2 (amorolfine[15]), and EBP (TASIN-1[16]) enhanced formation of MBP^+ oligodendrocytes, whereas inhibitors of the five other pathway enzymes were ineffective (Fig. 2b, Extended Data Fig. 3e–h). Treatments had little effect on cell number (Extended Data Fig. 3e). Concentrations of amorolfine and TASIN-1 that enhanced oligodendrocyte formation also led to accumulation of 14-dehydrozymostenol and zymostenol, respectively (Extended Data Fig. 3i, j). Moreover, distinct structural classes of inhibitors of CYP51, TM7SF2 and EBP comparably enhanced oligodendrocyte formation, including at picomolar doses[17] (Extended Data Fig. 4a–h).

We also used CRISPR–Cas9 targeting to evaluate the effects of genetic suppression of *EBP*. OPCs expressing Cas9 and guide RNA targeting *EBP* demonstrated reduced *EBP* transcript levels, robust accumulation of the expected intermediate zymostenol, and enhanced formation of oligodendrocytes under differentiation-permissive conditions (Fig. 2c, d, Extended Data Fig. 4k). Two independent guide RNA sequences produced similar results (Extended Data Fig. 4i–l). In total, this genetic and chemical genetic analysis suggests that inhibition of the cholesterol biosynthesis pathway within a limited window of enzymes between CYP51 and EBP is sufficient to enhance the formation of oligodendrocytes.

The efficacy of these small molecules and genetic perturbations was not mediated by a simple reduction in sterol levels, as treatment with statin drugs or methyl β-cyclodextrin depleted cholesterol levels comparably without enhancing oligodendrocyte formation (Fig. 2b, Extended Data Figs. 3a, b, 5a, b). Because treatment of OPCs with the

[1]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [2]Department of Pediatrics, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [3]Department of Anatomy and Regenerative Biology, George Washington University School of Medicine and Health Sciences, Washington, DC, USA. [4]Small Molecule Drug Development Core, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [5]Department of BioSciences, Rice University, Houston, TX, USA. [6]Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden, The Netherlands. [7]Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians University of Munich, Munich, Germany. [8]These authors contributed equally: Zita Hubler, Dharmaraja Allimuthu. *e-mail: drew.adams@case.edu

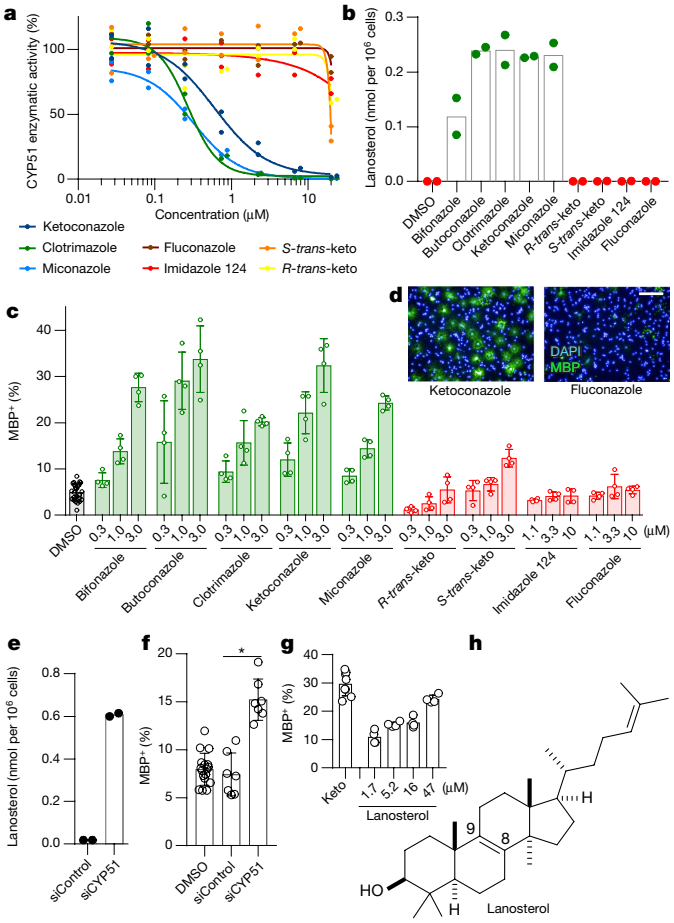


Fig. 1 | Imidazoles inhibit CYP51 to enhance oligodendrocyte formation. a, Rat CYP51 enzymatic activity following treatment with azoles. $n = 2$ independent enzymatic assays. **b**, GC–MS-based quantification of lanosterol levels in OPCs treated with the indicated azoles at 2.5 μM. $n = 2$ wells per condition. **c**, **f**, **g**, Percentage of MBP^+ oligodendrocytes generated from OPCs following treatment with azoles (**c**), cell-permeable siRNA reagents (**f**), or lanosterol (**g**). $n \geq 4$ wells per condition; for exact well counts in all figures, see Methods section 'Statistics and reproducibility'. In **f**, $^*P = 0.0005$, two-tailed Student's *t*-test. **d**, Representative images of OPCs treated with the indicated azoles. Nuclei are labelled with DAPI (blue) and oligodendrocytes are indicated by immunostaining for MBP (green). Scale bar, 100 μm. **e**, GC–MS-based quantification of lanosterol levels in OPCs treated with the indicated reagents. $n = 2$ wells per condition. **h**, Structure of lanosterol. All bar graphs indicate mean ± s.d. Results in **c**, **d**, **g** are representative of three independent experiments; those in **b**, **e**, **f** are representative of two independent experiments using OPC-5 cells; for validation in an independent derivation of OPCs, see Extended Data Fig. 2. Keto, ketoconazole.

CYP51 substrate lanosterol enhanced oligodendrocyte formation, we examined the effects of other purified sterols. Treatment of OPCs with 8,9-unsaturated sterols, including 14-dehydrozymostenol (which accumulates following TM7SF2 inhibition) and zymostenol (which accumulates following EBP inhibition), enhanced the formation of MBP^+ oligodendrocytes. By contrast, sterols lacking 8,9-unsaturation, including cholesterol itself[18], were ineffective (Fig. 2e, h, Extended Data Fig. 5c). A total of nine natural and unnatural 8,9-unsaturated sterols enhanced oligodendrocyte formation from OPCs, with 2,2-dimethylzymosterol the most potent among those evaluated to date (Fig. 2f, Extended Data Fig. 5d–l, o). Conversely, treating OPCs with Ro 48-8071, which inhibits lanosterol synthase and thereby prevents the accumulation of 8,9-unsaturated sterols, abrogated the enhanced oligodendrocyte formation induced by the CYP51 inhibitor ketoconazole (Extended Data Fig. 5m, n, p). In addition, analogues of either zymostenol or 8-dehydrocholesterol that lacked 8,9-unsaturation were inactive, demonstrating that 8,9-unsaturation is a crucial structural feature for activity in OPCs (Fig. 2g, Extended Data Fig. 5k, l). Finally, co-treating OPCs with ketoconazole and MAS-412 provided no further benefit over ketoconazole alone, confirming that these molecules act through a redundant mechanism (Extended Data Fig. 5q, r). Together these findings indicate that the accumulation of 8,9-unsaturated sterols in OPCs is a central mechanism for enhancing oligodendrocyte formation, whether these sterols arise from small-molecule inhibition of cholesterol biosynthesis enzymes or are supplied to OPCs in purified form.

Most of the 8,9-unsaturated sterols that are shown here to enhance oligodendrocyte formation have previously been shown to function as signalling lipids in oocytes by inducing the resumption of meiosis[19,20]. While the direct cellular targets of 8,9-unsaturated 'meiosis-activating sterols' remain poorly understood, there is evidence nuclear hormone receptors (NHRs) may play a role[19]. We evaluated 2,2-dimethylzymosterol and the pathway inhibitors ketoconazole and TASIN-1 in cell-based reporter assays for 20 NHRs, but no molecule showed significant activity in any assay (Extended Data Fig. 5s–u). Additional experiments discounted a role for SREBP2, which transcriptionally regulates cholesterol homeostasis, suggesting that these sterols act by mechanisms beyond NHRs or SREBP2 (Extended Data Fig. 5v). Together, these studies suggest a novel role for the meiosis-activating sterols in promoting oligodendrocyte formation.

In parallel, we executed a screen of over 3,000 bioactive small molecules and approved drugs at a uniform dose of 2 μM (Extended Data Fig. 6a). In addition to molecules previously annotated as enhancing OPC differentiation[5,6,9], we also identified many confirmed hits whose known targets did not cluster into easily discernible categories (Supplementary Table 1). Among the top ten novel enhancers of oligodendrocyte formation, four molecules had previously been shown to inhibit TM7SF2 or EBP in CNS-derived cells[11,21]. In fact, GC–MS-based sterol profiling revealed that all ten top hits led to accumulation of 8,9-unsaturated sterols at the screening dose, whereas randomly selected library members had no effect on sterol levels or oligodendrocyte formation. (Fig. 3a, Extended Data Fig. 6b–f).

Given the frequency of cholesterol pathway modulators within our top screening hits, we assessed whether any previously reported enhancers of remyelination identified by drug screening might also induce accumulation of sterol intermediates. At concentrations that promoted oligodendrocyte formation, benztropine, clemastine, tamoxifen, and U50488 induced accumulation of zymostenol and zymosterol and decreased basal sterol levels, indicative that they inhibited EBP in OPCs (Fig. 3b, Extended Data Fig. 6g–l). Tamoxifen has been shown to inhibit the enzymatic activity of EBP directly[11,22,23], and we confirmed that benztropine, clemastine, tamoxifen, U50488, and several high-throughput screening (HTS) hits all inhibited EBP directly in a biochemical assay[22] (Fig. 3c). By contrast, liothyronine and bexarotene showed minimal effects on sterol levels in OPCs (Fig. 3b, Extended Data Fig. 6g), consistent with their known functions as modulators of transcription factor function and confirming that many, but not all, treatments that enhance oligodendrocyte formation cause accumulation of 8,9-unsaturated sterols.

While each of these bioactive small molecules has a previously annotated 'canonical' target, extensive structure–activity relationship data show that the ability to inhibit EBP, rather than the canonical target, predicts enhanced oligodendrocyte formation. For example, we validated a panel of six muscarinic receptor antagonists that all showed near-complete inhibition of the M1, M3, and M5 muscarinic receptor subtypes at the HTS dose of 2 μM (Extended Data Fig. 6m, p). Among these molecules, only clemastine and benztropine inhibited EBP in OPCs, and only clemastine and benztropine enhanced oligodendrocyte formation (Extended Data Fig. 6j, k, m–r). Likewise, among selective oestrogen receptor modulators (SERMs), toremifene and ospemifene are structurally near-identical and show comparable cellular anti-oestrogen activity. However, only toremifene inhibited EBP in OPCs, and only toremifene enhanced oligodendrocyte formation

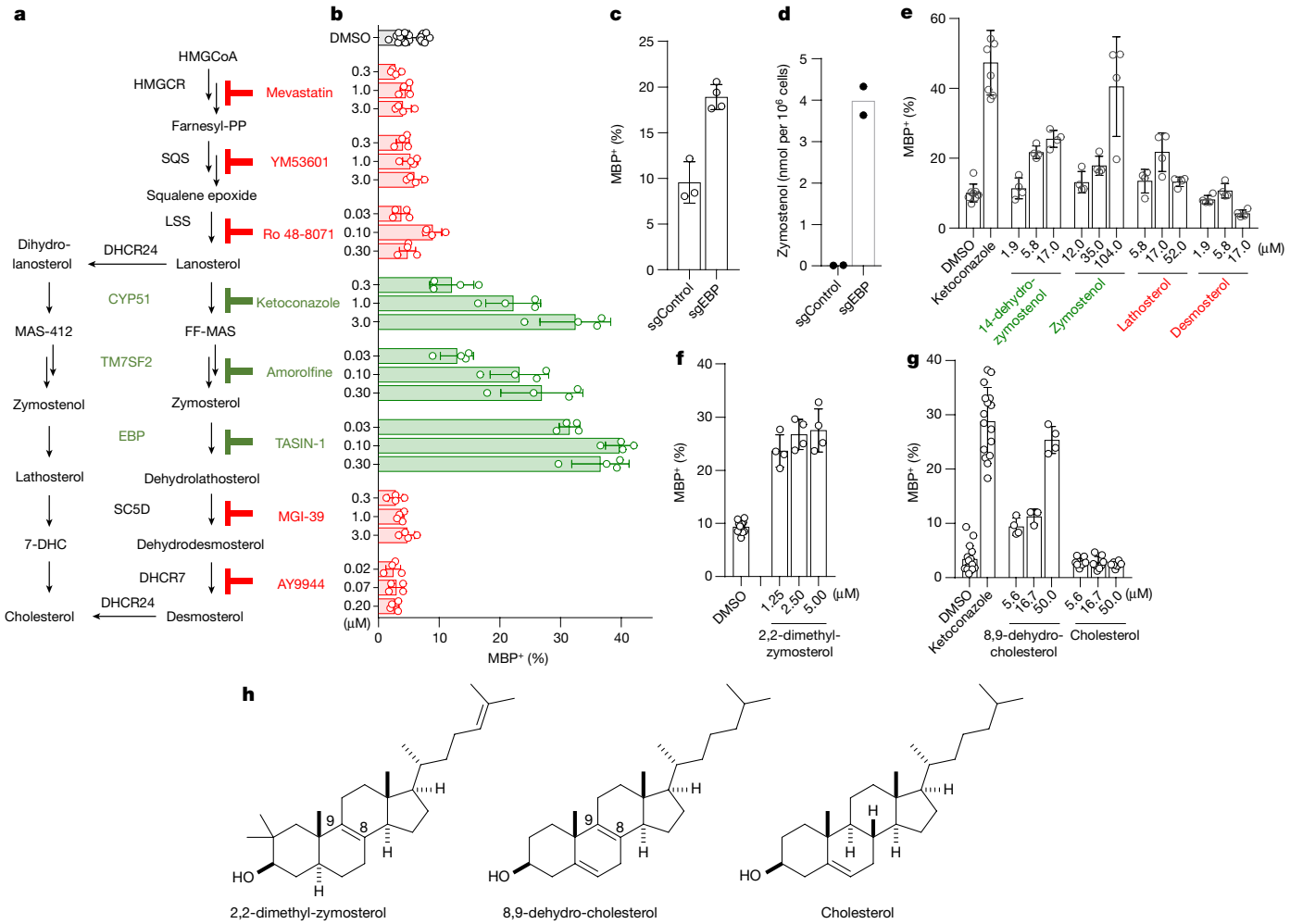


Fig. 2 | Small-molecule inhibition of CYP51, TM7SF2, or EBP enhances oligodendrocyte formation via accumulation of 8,9-unsaturated sterols. **a**, Abbreviated cholesterol biosynthesis pathway. For greater detail, see Extended Data Fig. 1. FF-MAS, follicular fluid-meiosis-activating sterol. **b**, Percentage of MBP$^+$ oligodendrocytes generated from OPCs treated with the indicated pathway inhibitors. $n \geq 4$ wells per condition. **c**, Percentage of MBP$^+$ oligodendrocytes generated from OPCs expressing Cas9 and guide RNA targeting EBP. $n \geq 3$ wells per condition. **d**, Functional validation of Cas9-based targeting of EBP using GC–MS-based quantification of zymostenol levels. $n = 2$ wells per condition. **e**–**g**, Percentage of MBP$^+$ oligodendrocytes generated from OPCs with the indicated purified sterols. $n \geq 4$ wells per condition. **h**, Structures of various sterols. All bar graphs indicate mean ± s.d. See Methods section 'Statistics and reproducibility' for exact well counts. Experiments in **b**–**g** are representative of two or more independent experiments using OPC-5 cells; for validation in an independent derivation of OPCs, see Extended Data Figs. 3–5.

(Extended Data Fig. 7a–g). Conversely, while 4-hydroxy-tamoxifen, as expected, showed 100-fold enhanced cellular anti-oestrogen activity relative to tamoxifen, both molecules have comparable potency for inhibition of EBP and comparable potency for enhancing oligodendrocyte formation (Extended Data Fig. 7h–j). Finally, the leading novel hit from our HTS, EPZ005687, was annotated as an inhibitor of the histone methyltransferase EZH2. However, analysis of three additional structurally related EZH2 inhibitors revealed that only EPZ005687 inhibited EBP and enhanced oligodendrocyte formation (Extended Data Fig. 7k–r). Across various antimuscarinic agents, SERMs, and EZH2 inhibitors, the ability to inhibit EBP, rather than each molecule's canonical activity, predicted enhanced oligodendrocyte formation.

We next tested the potential for combinations of small molecules to show additive or non-additive effects. Combining the thyroid hormone agonist liothyronine with a range of treatments that both modulated sterols and induced differentiation of OPCs produced additive effects on oligodendrocyte formation, indicating that these molecules are likely to function by mechanisms other than thyroid hormone receptor signalling to enhance oligodendrocyte generation (Extended Data Fig. 8a, b). By contrast, combinations of ketoconazole at a maximally effective dose with benztropine, clemastine, tamoxifen, or U50488 did not enhance differentiation above levels seen for ketoconazole alone (Extended Data Fig. 8c–e), consistent with these molecules sharing 8,9-unsaturated sterol accumulation as a common mechanism for induction of oligodendrocyte formation.

Because our in vitro OPC assays modelled only the initial differentiation event into oligodendrocytes, we next tested whether sterol pathway modulation also enhanced subsequent oligodendrocyte maturation and myelination in vitro and in vivo. First, we cultured OPCs on electrospun microfibres to assess the effects of sterol pathway modulators on the ability of oligodendrocytes to track and wrap along axon-like substrates[24,25]. Ketoconazole (CYP51), amorolfine (TM7SF2), and TASIN-1 (EBP) all robustly enhanced tracking along and wrapping around microfibres by MBP$^+$ oligodendrocytes. By contrast, inhibition of other enzymes up- or downstream in the pathway had little effect on oligodendrocyte maturation and ensheathment of microfibres (Extended Data Fig. 8f–k).

The imidazole antifungal miconazole, which targets CYP51, penetrates the mouse blood–brain barrier and enhances remyelination in mouse models of demyelination[4]. Here we evaluated brain-penetrant molecules with affinity for TM7SF2 (ifenprodil) and EBP (tamoxifen) using a well-established mouse model in which injection of lysolecithin is used to create focal lesions of demyelination in the dorsal column white matter of the adult spinal cord[26]. In vehicle-treated mice,

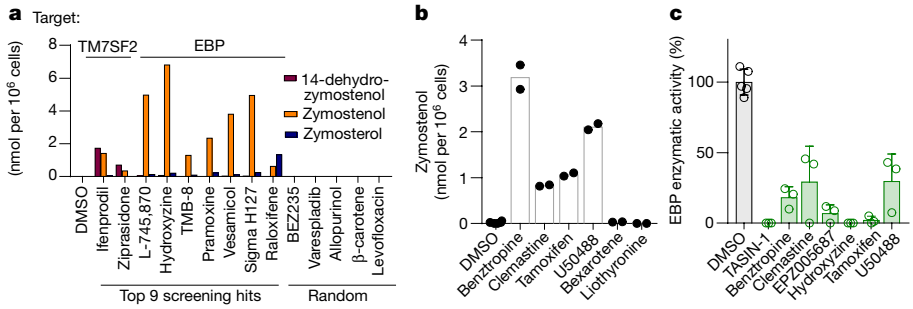


Fig. 3 | Inhibition of TM7SF2 and EBP is a unifying mechanism for many small-molecule enhancers of oligodendrocyte formation. **a**, Quantification of sterol levels in OPCs treated with the indicated molecules at 2 μM (one well per condition; for validation in a second derivation of OPCs, see Extended Data Fig. 6). **b**, Quantification of sterol levels in OPCs treated with the indicated previously reported enhancers of oligodendrocyte formation ($n = 2$ wells per condition except DMSO, $n = 6$). Representative of two independent experiments; for concentrations, see Extended Data Fig. 6g. **c**, Quantification of EBP enzymatic activity in a biochemical assay. All treatments 10 μM. $n = 3$ independent enzymatic assays, except DMSO, $n = 5$. Bars indicate mean; error bars indicate s.d. Sigma H127, *p*-fluorohexahydro-sila-difenidol.

profiles of sparsely distributed remyelinating axons characterized by thin myelin sheaths were detected mainly at the periphery of the lesion, while ultrastructural analyses revealed unmyelinated axons or axons with a single wrap of myelin (Fig. 4a, b). By contrast, after eight days of treatment with ifenprodil or tamoxifen, remyelination was widespread throughout the lesion (Fig. 4a, b, Extended Data Fig. 9a), consistent with a recent report regarding tamoxifen[9]. Critically, we used GCMS-based sterol profiling of brain tissue from mice treated with miconazole, ifenprodil, and tamoxifen to demonstrate that these therapeutic dosing regimens all led to substantial accumulation of 8,9-unsaturated sterols within the mouse brain, indicating that CYP51, TM7SF2, and EBP, respectively, were inhibited (Fig. 4c). Collectively, these data show that small-molecule inhibitors of CYP51, TM7SF2, and EBP can engage their sterol pathway targets and enhance remyelination in mice.

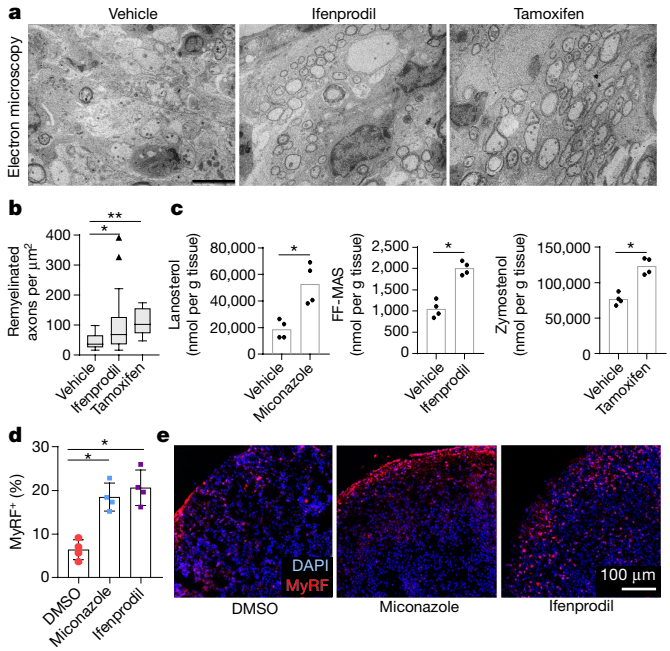


Fig. 4 | Accumulation of 8,9-unsaturated sterols enhances remyelination in vivo and in human brain spheroids. **a**, Representative electron microscopy images of LPC-lesioned dorsal spinal cord from mice treated with ifenprodil or tamoxifen. Scale bar, 5 μm. **b**, Tukey plot showing quantification of remyelinated axons in LPC-lesioned spinal cord from mice in **a**. $n = 6$ animals per group except vehicle, $n = 4$. $^{**}P = 0.0004$, $^{*}P = 0.048$, two-tailed Mann–Whitney test. Boxes indicate the interquartile range, horizontal lines represent the median, and whiskers represent the smaller of 1.5 times the interquartile range and the minimum–maximum range. **c**, Quantification of brain sterol levels in mice treated with miconazole, ifenprodil, or tamoxifen. $n = 4$ animals per group. $P = 0.0007$ for miconazole, $P = 0.0003$ for ifenprodil, $P = 0.0006$ for tamoxifen; two-tailed Student's *t*-test. **d**, Quantification of myelin regulatory factor $(MYRF)^+$ oligodendrocytes within human myelinating cortical spheroids following treatment with miconazole (2 μM) or ifenprodil (2 μM). $n = 4$ spheroids per treatment condition. $P = 0.0009$ for miconazole, $P = 0.0009$ for ifenprodil; two-tailed Student's *t*-test. **e**, Representative images of spheroids. $DAPI^+$ nuclei (blue) and $MYRF^+$ oligodendrocytes (red) are labelled. Scale bar, 100 μm. In **c**, **d**, bar graphs indicate mean and error bars indicate s.d.

Finally, the oligodendrocyte-enhancing and sterol-modulating activities of leading pathway inhibitors extend to human cells and tissue. Various small molecules caused accumulation of the expected 8,9-unsaturated sterol intermediates both in a human glioma cell line and in human pluripotent stem cell-derived cortical spheroids[27], confirming that these molecules similarly engage the sterol synthesis pathway in mouse and human cells and CNS tissue (Extended Data Fig. 9b, c). Critically, miconazole and ifenprodil also substantially enhanced the generation human oligodendrocytes in a 3D human pluripotent stem cell-derived cortical spheroid model, indicating conservation of function across species (Fig. 4d, e).

We have defined a dominant mechanism shared by many small-molecule enhancers of remyelination: elevation of levels of 8,9-unsaturated sterol intermediates by inhibition of a narrow range of cholesterol biosynthesis enzymes between CYP51 and EBP. We have identified 27 small molecules that both enhance oligodendrocyte formation and increase levels of 8,9-unsaturated sterol intermediates[11,21,23]. Mechanistically, several lines of evidence support a central signalling role for 8,9-unsaturated sterols in the observed enhanced oligodendrocyte formation, including the ability of nine independent 8,9-unsaturated sterols to enhance the formation of oligodendrocytes when supplied to OPCs (Extended Data Fig. 10).

Myelin is cholesterol-enriched, and past work has established that genetic or pharmacological treatments that inhibit early enzymes in cholesterol biosynthesis lead to hypomyelination in vivo[28–30]. Our work supports these observations, as inhibition of HMGCoA reductase or squalene synthase had neutral-to-negative effects on oligodendrocyte formation in our assays (Fig. 2b, Extended Data Fig. 3). These enzymes catalyse steps before the synthesis of the first sterol intermediate, so their inhibition prevents the synthesis of all cellular sterols. Our findings establish an alternate paradigm in which the cholesterol biosynthesis pathway can be leveraged to enhance the formation of new oligodendrocytes by targeting later steps whose inhibition does not cause net depletion of cellular sterols. Instead, acute inhibition of CYP51, TM7SF2, or EBP during OPC differentiation induces a 'sterol shift' in which a minority of cellular cholesterol is diverted to 8,9-unsaturated sterol intermediates that signal to enhance oligodendrocyte formation. Notably, we and others have independently shown that multiple molecules now annotated by us as enhancing 8,9-unsaturated sterol intermediate levels can regenerate functional myelin in vivo, as evidenced by reversal of paralysis in mice with MS-like disease[4–6].

Ultimately, our work demonstrates that modulating the sterol landscape in OPCs can enhance the formation of oligodendrocytes and points to new therapeutic targets, potent inhibitors for these targets, and metabolite-based biomarkers to accelerate the development of optimal remyelinating therapeutics.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0360-3.

Received: 10 May 2017; Accepted: 3 May 2018;
Published online 25 July 2018.

1. Goldman, S. A., Nedergaard, M. & Windrem, M. S. Glial progenitor cell-based treatment and modeling of neurological disease. *Science* **338**, 491–495 (2012).
2. Fancy, S. P. et al. Overcoming remyelination failure in multiple sclerosis and other myelin disorders. *Exp. Neurol.* **225**, 18–23 (2010).
3. Franklin, R. J. & Ffrench-Constant, C. Remyelination in the CNS: from biology to therapy. *Nat. Rev. Neurosci.* **9**, 839–855 (2008).
4. Najm, F. J. et al. Drug-based modulation of endogenous stem cells promotes functional remyelination *in vivo*. *Nature* **522**, 216–220 (2015).
5. Deshmukh, V. A. et al. A regenerative approach to the treatment of multiple sclerosis. *Nature* **502**, 327–332 (2013).
6. Mei, F. et al. Micropillar arrays as a high-throughput screening platform for therapeutics in multiple sclerosis. *Nat. Med.* **20**, 954–960 (2014).
7. Mei, F. et al. Identification of the kappa-opioid receptor as a therapeutic target for oligodendrocyte remyelination. *J. Neurosci.* **36**, 7925–7935 (2016).
8. Huang, J. K. et al. Retinoid X receptor gamma signaling accelerates CNS remyelination. *Nat. Neurosci.* **14**, 45–53 (2011).
9. Gonzalez, G. A. et al. Tamoxifen accelerates the repair of demyelinated lesions in the central nervous system. *Sci. Rep.* **6**, 31599 (2016).
10. Lariosa-Willingham, K. D. et al. A high throughput drug screening assay to identify compounds that promote oligodendrocyte differentiation using acutely dissociated and purified oligodendrocyte precursor cells. *BMC Res. Notes* **9**, 419 (2016).
11. Korade, Z. et al. The effect of small molecules on sterol homeostasis: measuring 7-dehydrocholesterol in DHCR7-deficient Neuro2a cells and human fibroblasts. *J. Med. Chem.* **59**, 1102–1115 (2016).
12. Giera, M., Müller, C. & Bracher, F. Analysis and experimental inhibition of distal cholesterol biosynthesis. *Chromatographia* **78**, 343–358 (2015).
13. Giera, M., Plössl, F. & Bracher, F. Fast and easy *in vitro* screening assay for cholesterol biosynthesis inhibitors in the post-squalene pathway. *Steroids* **72**, 633–642 (2007).
14. Mir, F. & Le Breton, G. C. A novel nuclear signaling pathway for thromboxane A2 receptors in oligodendrocytes: evidence for signaling compartmentalization during differentiation. *Mol. Cell. Biol.* **28**, 6329–6341 (2008).
15. Jachak, G. R. et al. Silicon incorporated morpholine antifungals: design, synthesis, and biological evaluation. *ACS Med. Chem. Lett.* **6**, 1111–1116 (2015).
16. Zhang, L. et al. Selective targeting of mutant adenomatous polyposis coli (APC) in colorectal cancer. *Sci. Transl. Med.* **8**, 361ra140 (2016).
17. DeBrabander, J., Shay, J. W., Wang, W., Nijhawan, D. & Theodoropoulos, P. Targeting emopamil binding protein (EBP) with small molecules that induce an abnormal feedback response by lowering endogenous cholesterol biosynthesis. US patent application US 2016/0313302 A1 (2016).
18. Saher, G. et al. Therapy of Pelizaeus-Merzbacher disease in mice by feeding a cholesterol-enriched diet. *Nat. Med.* **18**, 1130–1135 (2012).
19. Byskov, A. G., Andersen, C. Y. & Leonardsen, L. Role of meiosis activating sterols, MAS, in induced oocyte maturation. *Mol. Cell. Endocrinol.* **187**, 189–196 (2002).
20. Grøndahl, C. Oocyte maturation. Basic and clinical aspects of *in vitro* maturation (IVM) with special emphasis of the role of FF-MAS. *Dan. Med. Bull.* **55**, 1–16 (2008).
21. Canfrán-Duque, A. et al. Atypical antipsychotics alter cholesterol and fatty acid metabolism *in vitro*. *J. Lipid Res.* **54**, 310–324 (2013).
22. Moebius, F. F. et al. Pharmacological analysis of sterol delta8-delta7 isomerase proteins with [3H]ifenprodil. *Mol. Pharmacol.* **54**, 591–598 (1998).
23. Gylling, H. et al. Tamoxifen and toremifene lower serum cholesterol by inhibition of delta 8-cholesterol conversion to lathosterol in women with breast cancer. *J. Clin. Oncol.* **13**, 2900–2905 (1995).
24. Bechler, M. E., Byrne, L. & Ffrench-Constant, C. CNS myelin sheath lengths are an intrinsic property of oligodendrocytes. *Curr. Biol.* **25**, 2411–2416 (2015).
25. Lee, S. et al. A culture system to study oligodendrocyte myelination processes using engineered nanofibers. *Nat. Methods* **9**, 917–922 (2012).
26. Mi, S. et al. Promotion of central nervous system remyelination by induced differentiation of oligodendrocyte precursor cells. *Ann. Neurol.* **65**, 304–315 (2009).
27. Madhavan, M. et al. Induction of myelinating oligodendrocytes in human cortical spheroids. *Nat. Methods* https://doi.org/10.1038/s41592-018-0081-4 (2018).
28. Miron, V. E. et al. Statin therapy inhibits remyelination in the central nervous system. *Am. J. Pathol.* **174**, 1880–1890 (2009).
29. Klopfleisch, S. et al. Negative impact of statins on oligodendrocytes and myelin formation *in vitro* and *in vivo*. *J. Neurosci.* **28**, 13609–13614 (2008).
30. Saher, G. et al. High cholesterol level is essential for myelin membrane growth. *Nat. Neurosci.* **8**, 468–475 (2005).

Acknowledgements This work was supported by National Institutes of Health grant NS095280 (R.H.M., P.J.T), Conrad N. Hilton Foundation Pilot Innovator in MS Award (D.J.A.), Mallinckrodt Foundation Grant Award (D.J.A.), Mt. Sinai Health Care Foundation, philanthropic support from the Peterson, Fakhouri, Long, Goodman, Geller, Judge, and Weidenthal families, and unrestricted support from the CWRU School of Medicine. Z.H., M.S.E., K.C.A., Z.S.N., and J.L.S. were supported by the CWRU Medical Scientist Training Program (NIH T32 GM007250). Z.H. was also supported by NIH TL1 TR000441. Additional support was provided by the Small-Molecule Drug Development, Proteomics, and Translational Research Shared Resources of the Case Comprehensive Cancer Center (P30 CA043703). We acknowledge use of the Leica SP8 confocal microscope in the Light Microscopy Imaging Facility at CWRU made available through the Office of Research Infrastructure (NIH-ORIP) Shared Instrumentation Grant (S10 OD016164). We thank M. Drumm, T. Miller, B. Karl, O. Iyoha-Bello, J. Pink., P. Conrad, R. Lee, X. Li, D. Schlatzer, K. Polak, Janssen Pharmaceutica N.V., CXR Biosciences, ThermoFisher, Avanti Polar Lipids, and the P. Scacheri laboratory for technical assistance and discussion.

Author contributions Z.H., D.A., M.S.E., M.M., Z.S.N., K.C.A., H.E.S., M.A.T., and D.J.A. evaluated the effects of small molecules and genetic manipulations on oligodendrocyte formation in vitro. Z.H., D.A., I.B., M.A.T., F.B., and D.J.A. performed and analysed sterol profiling experiments in OPCs in vitro. D.C.F., Y.F., P.J.T., and D.J.A. performed high-throughput screening. Z.H., I.B., H.E.S., E.G., M.M., M.K., R.H.M., P.J.T., and D.J.A. evaluated the in vivo efficacy of small molecules on remyelination and sterol levels. Z.H. and J.L.S. profiled nuclear hormone receptors. Z.H., M.M., and Z.S.N. performed experiments on human cortical spheroids. J.J., W.K.W., M.G., and F.B. synthesized and purified sterol reagents. Z.H., D.A., P.J.T. and D.J.A. analysed all data and wrote the manuscript. All authors provided intellectual input, edited and approved the final manuscript.

Competing interests D.J.A., P.J.T., Z.H., D.A., M.S.E. and R.H.M. are inventors on patents and patent applications that relate to this work and have been licensed to Convelo Therapeutics, Inc., which seeks to develop remyelinating therapeutics. D.J.A. and P.J.T. hold equity in Convelo Therapeutics, Inc. and receive consulting income from Convelo Therapeutics, Inc. After resubmission of this work, D.C.F. became an employee of Convelo Therapeutics, Inc.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0360-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0360-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.J.A.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Statistics and reproducibility. No statistical methods were used to predetermine sample size. Data were expressed as mean ± s.d. and *P* values were calculated using an unpaired two-tailed Student's *t*-test for pairwise comparison of variables with a 95% confidence interval and $n-2$ degrees of freedom, where n is the total number of samples, in all figures except Fig. 4b. In Fig. 4b, *P* values were calculated using an unpaired two-tailed Mann–Whitney test with 95% confidence interval and the data plotted as a Tukey box and whisker plot. Boxes indicate the interquartile range, and the horizontal line represents the median. Biological replicates: Fig. 1c, $n=4$ wells per condition, except DMSO, $n=24$; Fig. 1f, $n=17$ wells for DMSO, $n=7$ for siControl and siCYP51; Fig. 1g, $n=8$ wells for DMSO and $n=4$ for lanosterol; Fig. 2b, $n=4$ wells per condition, except DMSO, $n=24$; Fig. 2c, $n=3$ wells for sgControl and $n=4$ for sgEBP; Fig. 2e–g, $n=4$ wells per condition, except $n=8$ for DMSO and $n=7$ for ketoconazole in Fig. 2e, $n=12$ for DMSO in Fig. 2f, $n=16$ for DMSO and ketoconazole, $n=8$ for cholesterol in Fig. 2g. Independent experiments: Fig. 2b, f are representative of three and Fig. 2c, e, g of two independent experiments using OPC-5 cells; for validation in an independent derivation of OPCs, see Extended Data Figs. 3–5.

Small molecules. The identity and purity of small molecules were authenticated by LC–MS before use (Supplementary Table 2). The following compounds were purchased from Sigma-Aldrich as solids: ketoconazole, miconazole, clotrimazole, fluconazole, fulvestrant, ifenprodil, benztropine, bexarotene, tamoxifen, 4-hydroxytamoxifen, medroxyprogesterone acetate, ospemifene, GSK343, *trans*-u50488, methyl-β-cyclodextrin, 5α−cholestan-3β-ol, and cholesterol. The following compounds were purchased from Cayman Chemicals as solids: liothyronine, clemastine, AY9944, YM53601 and Ro-48-8071. The following compounds were obtained from Janssen Pharmaceuticals as solids: 2-methyl-ketoconazole, *R-trans*-ketoconazole, and *S-trans*-ketoconazole. Mevastatin was purchased as a solid from Selleck Chemicals. The following compounds were purchased from Selleck Chemicals as 10 mM DMSO solutions: bifonazole, butoconazole, amorolfine, toremifene, EPZ005687, EPZ6438, UNC1999, hydroxyzine, ziprasidone, *p*-fluorohexahydro-sila-difenidol (abbreviated in figures as Sigma H127), vesamicol, raloxifene, L-745,870, TMB-8, pramoxine, varespladib, tanshinone-I, levofloxacin, nateglinide, abiraterone, allopurinol, detomidine, rivastigmine, β-carotene, BEZ-235, scopolamine, and homatropine. Pirenzepine and telenzepine were purchased from Sigma-Aldrich as 10 mM DMSO solutions. Cholesterol biosynthetic intermediates were purchased from Avanti Polar Lipids as solids: lanosterol, zymosterol, zymostenol, lathosterol, desmosterol, 7-dehydrodesmosterol, FF-MAS (4,4-dimethyl-5α-cholesta-8,14,24-trien-3β-ol), 8,9-dehydrocholesterol, and 2,2-dimethylzymosterol (2,2-dimethyl-5α-cholesta-8,24-dien-3β-ol). 14-Dehydrozymostenol (5α-cholesta-8,14-dien-3β-ol), MAS-412 (4,4-dimethyl-5α-cholesta-8,14-dien-3β-ol), and MAS-414 (4,4-dimethyl-5α-cholesta-8-en-3β-ol) were provided by F.B. Imidazole 124[31], TASIN-1[16], TASIN-449[17], and MGI39[32] were synthesized as reported. T-MAS (4,4-dimethyl-5α-cholesta-8,24-dien-3β-ol) from HPLC purification of yeast extracts was provided by J.J. and W.K.W.

Mouse OPC preparation. To rigorously assess the effects of small-molecule and genetic treatments on OPCs, all treatments were assayed in two batches of epiblast stem cell-derived OPCs, and key results were confirmed using mouse primary OPCs. OPCs were generated from two separate EpiSC lines, EpiSC5 (giving rise to OPC-5 OPCs) and 129O1 (giving rise to OPC-1 OPCs). Unless otherwise noted, results in OPC-5 cells are presented in Figs. 1–4 while results in OPC-1 are presented in Extended Data Figs. 1–10.

EpiSC-derived OPCs were obtained using in vitro differentiation protocols and culture conditions described previously[33]. To ensure uniformity throughout all in vitro screening experiments, EpiSC-derived OPCs were sorted to purity by fluorescence activated cell sorting at passage five with conjugated CD 140a-APC (eBioscience, 17-1401; 1:80) and NG2-AF488 (Millipore, AB5320A4; 1:100) antibodies. Sorted batches of OPCs were expanded and frozen down in aliquots. OPCs were thawed into growth conditions for one passage before use in further assays. Cultures were regularly tested and shown to be mycoplasma free.

To obtain mouse primary OPCs, whole brain was removed from post-natal day 2 pups anaesthetized on ice. Brains were placed in cold DMEM/F12, and the cortices were isolated and the meninges were removed. The cortices were manually chopped and processed with the Tumour Dissociation Kit (Miltenyi) and incubated at 37 °C for 10 min. The cell suspension was filtered through a 70 μm filter and centrifuged at 200*g* for 4 min at room temperature. The cells were washed in DMEM/F12, re-centrifuged and plated in poly-Ornithine and Laminin-treated flasks containing DMEM/F12 supplemented with N2 Max, B27 (ThermoFisher), 20ng/ml FGF, and 20ng/ml PDGF. OPCs were passaged once before treatment. Media was changed every 48 h.

In vitro phenotypic screening of OPCs. EpiSC-derived OPCs were grown and expanded in poly-ornithine (PO) and laminin-coated flasks with growth medium (DMEM/F12 supplemented with N2-MAX (R&D Systems), B-27 (ThermoFisher), GlutaMax (Gibco), FGF2 (10 μg/ml, R&D systems, 233-FB-025) and PDGF-AA (10 μg/ml, R&D systems, 233-AA-050) before harvesting for plating. The cells were seeded onto poly-D-lysine 96-well CellCarrier or CellCarrierUltra plates (PerkinElmer) coated with laminin (Sigma, L2020; 15 μg/ml) using multi-channel pipet. For the experiment, 800,000 cells/ml stock in differentiation medium (DMEM/F12 supplemented with N2-MAX and B-27) was prepared and stored on ice for 2 h. Then, 40,000 cells were seeded per well in differentiation medium and allowed to attach for 30 min before addition of drug. For dose–response testing of all molecules except sterols, a 1,000× compound stock in dimethyl sulfoxide (DMSO) was added to assay plates with 0.1 μl solid pin multi-blot replicators (V & P Scientific; VP 409), resulting in a final primary screening concentration of 1×. Sterols were added to cells as an ethanol solution (0.2% final ethanol concentration). Positive control wells (ketoconazole, 2.5 μM) and DMSO vehicle controls were included in each assay plate. Cells were incubated under standard conditions (37 °C, 5% CO_2) for 3 days and fixed with 4% paraformaldehyde (PFA) in phosphate buffered saline (PBS) for 20 min. Fixed plates were washed with PBS (200 μl per well) twice, permeabilized with 0.1% Triton X-100 and blocked with 10% donkey serum (v/v) in PBS for 40 min. Then, cells were labelled with antibodies recognizing MBP (Abcam, ab7349; 1:200) or PLP1 (1:1,000, clone AA3, generously provided by B. Trapp, Cleveland Clinic) for 16 h at 4 °C followed by detection with Alexa Fluor conjugated secondary antibodies (1:500) for 45 min. Nuclei were visualized by DAPI staining (Sigma; 1 μg/ml). During washing steps, PBS was added using a multi-channel pipet and aspiration was performed using Biotek EL406 washer dispenser (Biotek) equipped with a 96-well aspiration manifold.

High-content imaging and analysis. Plates were imaged on the Operetta High Content Imaging and Analysis system (PerkinElmer) and a set of 6 fields captured from each well resulting in an average of 1,200 cells being scored per well. Analysis (PerkinElmer Harmony and Columbus software) began by identifying intact nuclei stained by DAPI; that is, those traced nuclei that were larger than 300 μm^2 in surface area. Each traced nucleus region was then expanded by 50% and cross-referenced with the mature MBP stain to identify oligodendrocyte nuclei, and from this the percentage of oligodendrocytes was calculated. In some experiments, PLP1 staining was performed instead of MBP, or the total process length of MBP^+ oligodendrocytes was calculated as previously described[4].

OPCs differentiation and sterol profiling after methyl-β-cyclodextrin treatment. EpiSCs derived OPCs harvested from culture flasks were resuspended in 10 ml of differentiation medium to a final cell density of 500,000 cells/ml. To this, cell-culture grade water or methyl-β-cyclodextrin (1 mM) was added and incubated at 37 °C. After 30 min the cells were washed twice with differentiation medium (5 ml), and split into two portions for differentiation and sterol profiling. The 1,000,000 cells per condition were directly processed as described in GC–MS-based sterol profiling to measure the endogenous sterol levels. For differentiation, the cells were resuspended in differentiation medium to a final cell density of 800,000 cells/ml and plated in a PDL/laminin coated 96-well CellCarrierUltra plate. After 72 h, the cells were fixed, stained, imaged and quantified as described above.

High-throughput screening of 3,000 bioactive small molecules. EpiSC-derived OPCs were grown and expanded in poly-ornithine and laminin-coated flasks before harvesting for plating. Cells were dispensed in differentiation medium supplemented with Noggin (R&D Systems; 100 ng/ml), Neurotrophin 3 (R&D Systems; 10 ng/ml), cAMP (Sigma; 50 μM), and IGF-1 (R&D Systems; 100 ng/ml)) using a Biotek EL406 Microplate Washer Dispenser (Biotek) equipped with 5 μl dispense cassette (Biotek), into poly-D-lysine/laminin (Sigma, L2020; 4 μg/ml)-coated sterile, 384-well, CellCarrier ultra plates (PerkinElmer), to a final density of 12,500 cells per well and allowed to attach for 45 min before addition of drug. A 3 mM stock of bioactive compound library in dimethylsulphoxide (DMSO) were prepared in an Abgene storage 384-well plate (ThermoFisher Scientific; AB1055). These were added to assay plates using a 50 nl solid pin tool attached to a Janus automated workstation (Perkin Elmer), resulting in a final screening concentration of 2 μM. Cells were incubated at 37 °C for 1 h and then T3 (Sigma; 40 ng/ml) was added to all wells except negative controls, to which FGF (20 ng/ml) was added instead. Negative controls and T3-alone were included in each assay plate. After incubation at 37 °C for 72 h, cells were fixed, washed and stained similarly to the 96-well OPC assay protocol, although all the washing steps were performed using a Biotek EL406 Microplate Washer Dispenser (Biotek) equipped with a 96-well aspiration manifold. Cells were stained with DAPI (Sigma; 1 μg/ml) and MBP antibody (Abcam, ab7349; 1:100). Plates were imaged on the Operetta High Content Imaging and Analysis system (PerkinElmer) and a set of 4 fields captured from each well resulting in an average of 700 cells being scored per well. Analysis was performed as in High-Content Imaging and Analysis, above. All plates for the primary screen were processed and analysed simultaneously to minimize variability. Molecules causing more than 20% reduction in nuclear count relative to DMSO control wells were removed from consideration, and hits were called on the basis of largest fold-increase in percentage of MBP^+ oligodendrocytes relative to DMSO controls within the same plate. When selecting the leading hits for

further experiments, molecules obtained in previous screens were omitted, including imidazole antifungals and clemastine.

GC–MS-based sterol profiling. EpiSC-derived OPCs were plated at 0.5 million cells per ml in PDL- and laminin-coated six or twelve well plate with differentiation media. After 24 h, cells were dissociated with Accutase, rinsed with saline, and cell pellets were frozen. For sterol analyses, cells were lysed in methanol (Sigma-Aldrich) with agitation for 30 min and cell debris removed by centrifugation at 10,000 rpm for 15 min. Cholesterol-d7 standard (25,26,26,26,27,27,27-$^{2}H_{7}$-cholesterol, Cambridge Isotope Laboratories) was added before drying under nitrogen stream and derivatization with 55 μl of bis(trimethylsilyl)trifluoroacetamide/trimethylchlorosilane to form trimethylsilyl derivatives. Following derivatization at 60 °C for 20 min, 1 μl was analysed by GC–MS using an Agilent 5973 Network Mass Selective Detector equipped with a 6890 gas chromatograph system and a HP-5MS capillary column (60 m × 0.25 mm × 0.25 μm). Samples were injected in splitless mode and analysed using electron impact ionization. Ion fragment peaks were integrated to calculate sterol abundance, and quantitation was relative to cholesterol-d7. The following *m/z* ion fragments were used to quantitate each metabolite: cholesterol-d7 (465), FF-Mas (482), cholesterol (368), zymostenol (458), zymosterol (456), desmosterol (456, 343), 7-dehydrocholesterol (456, 325), lanosterol (393), lathosterol (458), 14-dehydrozymostenol (456). Calibration curves were generated by injecting varying concentrations of sterol standards and maintaining a fixed amount of cholesterol-d7. The human glioma cell line GBM528 was a gift of Jeremy Rich (Cleveland Clinic). These cells were validated as unique by STR profiling[34].

LC–MS-based sterol profiling. Sterols were extracted after treatment of OPC-5 OPCs with ketoconazole as described in GC–MS-based sterol profiling above. Picolinate derivatization, chromatographic separation, and mass spectrometric detection were performed as reported previously[35]. Peaks from selective reaction monitoring were integrated to calculate sterol abundance, and quantitation was relative to cholesterol-d7.

Human cortical spheroids. Human cortical spheroids were generated as described previously with modifications to enable the inclusion and differentiation of OPCs[27]. In brief, spheroids were treated with miconazole or ifenprodil (2 μM) from days 62-72 and assayed on day 93 for MyRF^{+} oligodendrocytes (rabbit anti-MyRF antibody was generously provided by M. Wegner and used at 1:1,000).

CYP51 enzymatic assay. CYP51 enzymatic activity was measured using a reported method with slight modifications[36]: rat CYP51 (Cypex, Inc.) was used as enzyme; reaction volume was 500 μl; reaction time was 30 min; lanosterol concentration was 50 μM; and reactions were quenched with 500 μl isopropanol. Finally, 15 μl of each reaction/isopropanol mixture was injected onto a SCIEX Triple Quad 6500 LC-MS/MS system using an APCI ion source in positive ion mode with a Shimadzu UFLC-20AD HPLC and a Phenomenex Kinetix C18XB 50 × 2.1 × 2.6 column at 40 °C.

EBP enzymatic assay. EBP enzymatic activity was measured using a reported method with slight modifications[22]: active EBP was obtained from mouse microsomes, inhibitors were added, zymostenol was added at a final concentration of 25 μM in a final reaction volume of 500 μl, and the reaction incubated at 37 °C for 2 h. Sterols were extracted using 3 × 1 ml hexanes, cholesterol-d7 was added to enable quantitation, and the pooled organics were dried (Na_2SO_4) and evaporated under nitrogen gas. Samples were then silylated and analysed using GC/MS as described above.

siRNA treatments. Cell-permeable siRNAs were obtained as pools of 4 individual siRNAs targeting mouse CYP51, or a non-targeting control (Accell siRNAs, Dharmacon. Pooled CYP51 siRNA sequence: GUCUGUUUUGAGAUUAGU; CGACUAUGCUUCGUUUAUA; CGCUGCUCUUCAAUAGUAA; CUAUUAAG UUAUUGUGAAC. Non-targeting control siRNA: UGGUUUACAUGUCG ACUAA). For differentiation analysis, cells were plated in a 96-well plate (as detailed above) and treated with 1 μM pooled siRNA suspended in RNase free water diluted in differentiation media (as detailed above). For sterol analysis cells were plated in a six-well plate at 300,000 cells per well in standard differentiation media supplemented with PDGF (R&D Systems, 20 ng/ml), neurotrophin 3 (R&D Systems; 10 ng/ml), cAMP (Sigma; 50 μM), IGF-1 (R&D Systems; 100 ng/ml), noggin (R&D Systems; 100 ng/ml). At 24 h, 1 μM siRNA was added to the media. Cells were grown for three more days in siRNA containing media, with growth factor supplementation every 48h, before harvesting and processing for GC–MS analysis as detailed above.

CRISPR–Cas9-mediated targeting of EBP. Guide RNA sequences were obtained using the Broad Brie library and manufactured by IDT. Nucleotide sequences (sgRNA sequence: GAAACGCAATCACTACCCAT (sgEBP); GGGGCCTAATTGTGATCACG (sgEBP2)) were prepared and inserted into the LentiCRISPRv2 plasmid (Addgene, 52961) using the instructions from GeckoLibrary preparation: in brief, Fastdigest BsbmB1 (fermentas) was used for plasmid digestion, T4 PNK (NEB M0201S) for nucleotide annealing, and Quick Ligase (NEB M2200S) for sgRNA insertion. Insertion was confirmed by Sanger sequencing. Hek293T cells were transfected using Lenti-x shots as per the manufacturer's protocol (Clontech). After 24 h the media was changed to OPC media for collection of virus. 48 h later the media was collected, supplemented with FGF, PDGF, and protamine sulfate (Sigma, 8 μg/ml), and used to transduce OPCs. 24 h later the media was changed to non-virus containing media for 48 h. Cells underwent two 48 h stretches of puromycin selection (Invitrogen). After 24 h of recovery in non-selection media, cells were plated for differentiation, GC–MS, and qPCR as described above.

Focal demyelination, drug treatment and histological analysis. Focal demyelination in the dorsal column of the spinal cord was induced by the injection of 1% LPC solution. 12 week old C57BL/6 female mice were anaesthetized using isoflurane and T10 laminectomies were performed. 1 μl of 1% LPC was infused into the dorsal column at a rate of 15 μl/h. At day 4, animals were randomized into treatment groups before treatment (2 animals were excluded due to surgical complications). Between days 4 and 11 post-laminectomy, animals received daily injections of either vehicle or drug intraperitoneally. Drugs were dissolved in DMSO or corn oil and then diluted with sterile saline for injections such that final doses were 2 mg/kg for tamoxifen and 10 mg/kg for ifenprodil. This experiment was done in a blinded manner: compounds were coded to ensure the researchers performing the experiments were unaware of the treatment being administered to each animal. All animals were euthanized 12 days post-laminectomy ($n = 4$–6 per group). Mice were anaesthetized using ketamine/xylazine rodent cocktail and then euthanized by transcardial perfusion with 4% PFA, 2% glutaraldehyde, and 0.1 M sodium cacodylate. Samples were osmicated, stained en bloc with uranyl acetate and embedded in EMbed 812, an Epon-812 substitute (EMS). 1 μm sections were cut and stained with toluidine blue and visualized on a light microscope (Leica DM5500B). The number of myelinated axons per unit area was counted from sections obtained from the middle of each lesion and then averaged over each treatment group. All sections within the lesion area were scored (vehicle, 10 sections; tamoxifen, 11 sections; ifenprodil, 28 sections). A Mann–Whitney statistical analysis was performed to assess statistical significance.

Analysis of mouse brain sterol levels. Ten to twelve week old male C57BL/6 mice were injected with 2mg/kg tamoxifen, 10 mg/kg ifenprodil, or 10 mg/kg miconazole dissolved in corn oil (tamoxifen) or DMSO (ifenprodil, miconazole) in sterile saline daily for three days. Mice were anaesthetized with isoflurane and perfused with phosphate buffered saline to remove blood from the brain. Brains were collected and flash frozen using liquid nitrogen. The samples were pulverized and 50-100 mg of tissue were collected for further processing. A modified Folch protocol was used for extraction of sterols[37]. Briefly, samples were resuspended in a 2:1 chloroform/methanol mixture and homogenized. Cell debris was removed by centrifugation at 4,000*g* for 10 min. The solution was dried under air and resuspended in hexane with a cholesterol-d7 standard and dried again. Lipids were derivatized with 70 μl of bis(trimethylsilyl)trifluoroacetamide; 2 μl were injected and analysed by GC–MS as described above.

Oestrogen-dependent cell proliferation assay. Oestrogen-dependent cell proliferation was measured as previously described with minor modifications[38]. After growth in oestrogen-free media (Phenol red-free RPMI supplemented with 10% charcoal stripped fetal bovine serum) for 5 days, cells were seeded at 2,500 cells/well into 96 well plates. The following day 3× drug containing media was added to triplicate wells and cells were allowed to grow for an additional 5 days at 37 °C in a standard 5% CO_2 humidified incubator. Total DNA per well was measured using an adaptation of the method of Labarca and Paigen[39]. At this time media was removed, cells were washed one time with 0.25× PBS and 100 μl of distilled water was added. Plates were frozen and thawed to enhance cell lysis and 200 μl of 10 μg/ml Hoechst 33258 (Sigma-Aldrich, St. Louis, MO.) in 2 M NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 7.4 was added. After incubation at room temperature for 2 h, plates were read in a SpectraMax i3 fluorescent plate reader (Molecular Devices, Sunnyvale, CA) with excitation at 360 nm and emission at 460 nm. All values were converted to microgram DNA per well using a standard curve derived from purified salmon testes DNA. T47D cells were provided by the Translational Research Shared Resource of the Case Comprehensive Cancer Center and used without further authentication beyond the observed oestrogen-dependent cell proliferation.

Oligodendrocyte formation and imaging on electrospun microfibres. A 12-well plate containing Mimetex aligned scaffold (microfibre plate, AMSBIO, AMS.TECL-006-1X, Electrospun poly-L-lactide Scaffold, 2 μM fibre diameter cell crown inserts) was prepared as previously described[24]. In brief, fibre inserts were sterilized with 70% ethanol and washed with PBS before being coated with poly-ornithine and laminin. After laminin coating, 100,000 cells/ml of EpiSC-derived OPCs (1.5 ml/well) were plated in differentiation medium. After 24 h the media was replaced with fresh media containing small-molecule treatments. Every 48 h the media was replaced with fresh compound containing media for a total of 14 days. Plates were fixed with 4% PFA, permeabilized with 0.1% Triton X-100, and blocked with 10% donkey serum (v/v) in PBS for 60 min. Plates were stained for MBP (Abcam, ab7349; 1:100) and DAPI staining (Sigma; 5 μg/ml). After staining, the inserts were moved into new 12-well plate and covered with 2 ml of PBS before imaging in Operetta high content Imaging and analysis system. Plates were imaged

on the Operetta High Content Imaging and Analysis system (PerkinElmer) and a set of 8 fields captured from each well resulting in an average of 45,000 cells being scored per well. Analysis (PerkinElmer Harmony and Columbus software) identified intact nuclei stained by DAPI and calculated the MBP signal intensity per cell per well. Microfibre insert tracking images were taken using a Leica DMi8 with a 20× dry/NA 0.40 objective. Microfibre plate inserts were mounted using Flouromount-G (SouthernBiotech) and allowed to partially harden before coverslips were added and the insert ring was removed. Confocal images were obtained on a Leica SP8 confocal scanning microscope, with 40× oil/NA 1.30 objective. Confocal stacks of 0.336 μm *z*-steps were taken at 1,024 × 1,024. Each fluorophore was excited sequentially and all contrast and brightness changes were applied consistently between images.

A separate analysis approach was performed on an independent experiment performed as above except the small-molecule treatment was limited to the first 4 days of the 14 day culture period. After staining, the fibre inserts were mounted on a glass slide (Fisherbrand Superfrost Plus Microscope Slides) using Fluormount-G (Southern Biotech) with a cover glass (Fisherbrand Microscope Cover Glass) and dried at RT in dark for 36 h. The mounted inserts were imaged on the Operetta High Content Imaging and Analysis system (PerkinElmer) and a set of 22 fields captured from each condition resulting in an average of 2,000 cells being scored per well. The total microfibre area was calculated using bright filed imaging and a spot-finding function (area larger than 2 pixels^2). The MBP+ pixel area within the defined microfibre area was then defined and the percentage of the total microfibre area calculated.

CYP51 qPCR. Cells were plated at 500,000 cells per well in a six-well plate and were grown in standard differentiation media supplemented with PDGF, neurotrophin 3, cAMP, IGF-1, and noggin for four days as described above. At 24 h, cells were treated with 1 μM siRNA. Growth factors were added every 48 h. After three days of siRNA treatment, RNA was isolated with the RNeasy Mini Kit (Qiagen), and cDNA was made using High-Capacity RNA-to-cDNA Kit (Applied Biosystems). Exon spanning primers for ActinB (Thermo-Fisher, Taqman, Mm02619580_g1) and CYP51 (Thermo-Fisher, Taqman, Mm00490968_m1) were used for detection of relative RNA levels by quantitative real time PCR (Applied Biosystems, 7300 Realtime PCR system). Cycle time and outliers were calculated using Applied Biosystems' 7300 System Sequence Detection Software version 1.4.

EBP qPCR. OPCs were accutased, 1 million cells per cell line were spun down and RNA was isolated with the RNeasy Mini Kit (Qiagen). DNA was removed using DNAse (Invitrogen), and cDNA was made using High-Capacity RNA-to-cDNA Kit (Applied Biosystems). Primers for exon 5 of EBP (forward primer: TGTGC GAGGAGGAAGAAGAT, reverse primer: GATAGGCCACCCCGTTTATT) and GAPDH (forward primer: AGGTCGGTGTGAACGGATTTG; reverse primer: GGGGTCGTTGATGGCAACA) were manufactured by IDT and gene expression was assessed using Power SYBR Green Master Mix (Applied Biosystems) were used for detection of relative RNA levels by quantitative real time PCR (Quantstudio 7 flex system). Cycle time and outliers were calculated using QuantStudio Software V1.3.

Muscarinic receptor antagonism assay. GeneBLAzer M1-NFAT-*bla* CHO-K1 cells (or M3- or M5-NFAT-*bla* CHO-K1 cells) (ThermoFisher) were thawed into Assay Media (DMEM, 10% dialysed FBS, 25 mM HEPES pH 7.3, 0.1 mM NEAA). 10,000 cells/well were added to a 384-well TC treated assay plate and incubated 16–24 h at 37 °C. 4 μl of a 10× stock of antimuscarinic molecules was added to the plate and incubated 30 min. 4 μl of 10× control agonist Carbachol at the predetermined EC80 concentration was added to wells containing antimuscarinic molecules. The plate was incubated 5 h and 8 μl of 1 μM substrate + solution D loading solution was added to each well and the plate was incubated 2 h at room temperature before reading on a fluorescence plate reader. This cell line was validated in each run on the basis of $z' > 0.5$ for carbachol versus control treatment.

SREBP qPCR. Cells were plated at 1 million cells per well in a six-well plate and were grown in standard differentiation media supplemented with with DMSO, mevastatin (2.5 μM), Ro 48-8071 (500 nM), ketoconazole (2.5 μM), TASIN-1 (100 nM), or amorolfine (100 nM). At 24 h, RNA was isolated with the RNeasy Mini Kit (Qiagen), and cDNA was made using High-Capacity RNA-to-cDNA Kit (Applied Biosystems). Exon spanning primers ActinB (Thermo-Fisher, Taqman, Mm02619580_g1), LSS (Thermo-Fisher, Taqman, Mm00461312_m1), LDLR (Thermo-Fisher, Taqman, Mm01177349_m1), and DHCR7 (Thermo-Fisher, Taqman, Mm00514571_m1) were used for detection of relative RNA levels by quantitative real time PCR (Applied Biosystems, 7300 Realtime PCR system). Cycle time and outliers were calculated using Applied Biosystems 7300 System Sequence Detection Software version 1.4.

NR2C2 and NR2F1 luciferase assays. Forty-eight hours before transfection, 100,000 Hek293T cells were plated per well in a 24 well plate. HEK293T cells were chosen because they were used previously in this assay and not validated further[40]. NR2C2 (Origene, MR221079) or NR2F1 (gift from C. Schaaf) and NGFI promoter reporter plasmid (gift from C. Schaaf) were transfected using Lipofectamine 2000 (Thermo-Fisher, 11668027) as per the manufacturer's protocol. After 16 h, Hek293 cells were treated with the compounds (2,2-dimethyl-zymosterol 5 μM, FF-MAS 10 μM, ketoconazole 2.5 μM, TASIN-1 100 nM, mevastatin 2.5 μM, liothyronine 3 μM, and all-*trans* retinoic acid 5 μM). 32 h later cells were lysed using a firefly luciferase assay system (Promega, E1500) and readout using Synergy Neo2 High Performance plate reader.

Nuclear receptor profiling. Luciferase reporter assays performed by Indigo Biosciences were used to assess interaction of 2,2-dimethylzymosterol (5 μM), ketoconazole (2.5 μM), and TASIN-1 (250 nM) with human ERα, GR, LXRβ, NFκB, NRF2, PGR, PPARα, PPARγ, RARα, RARγ, RXRα, RXRβ, TRα, TRβ and VDR in agonist mode and ERRα, RORα and RORγ in inverse-agonist mode. The reporter for these assays is firefly luciferase linked with either the genetic response elements (GRE) or the Gal4 upstream activation sequence (UAS). These cells also express either the native receptor or a receptor in which the native N-terminal DNA binding domain (DBD) has been replaced with that of the yeast Gal4 DBD. The specifics of each assay are shown in the table below. In brief, a suspension of reporter cells was prepared in cell recovery medium (CRM; containing 5% (RORγ) or 10% charcoal stripped FBS for others). 100 μl of the reporter cell suspension was dispensed into wells of a white 96-well assay plate. Test compound, reference compounds, and the respective vehicle were diluted into INDIGO's compound screening medium (CSM; containing 5% (RORγ) or 10% charcoal stripped FBS for others). 100 μl of each treatment medium was dispensed into duplicate assay wells pre-dispensed with reporter cells. Assay plates were incubated at 37 °C for 24 h. Following the incubation period, for agonist and inverse-agonist assays, treatment media were discarded and 100 μl/well of luciferase detection reagent was added. RLUs were quantified from each assay well to determine agonist or inverse-agonist activity using the following assay designs:

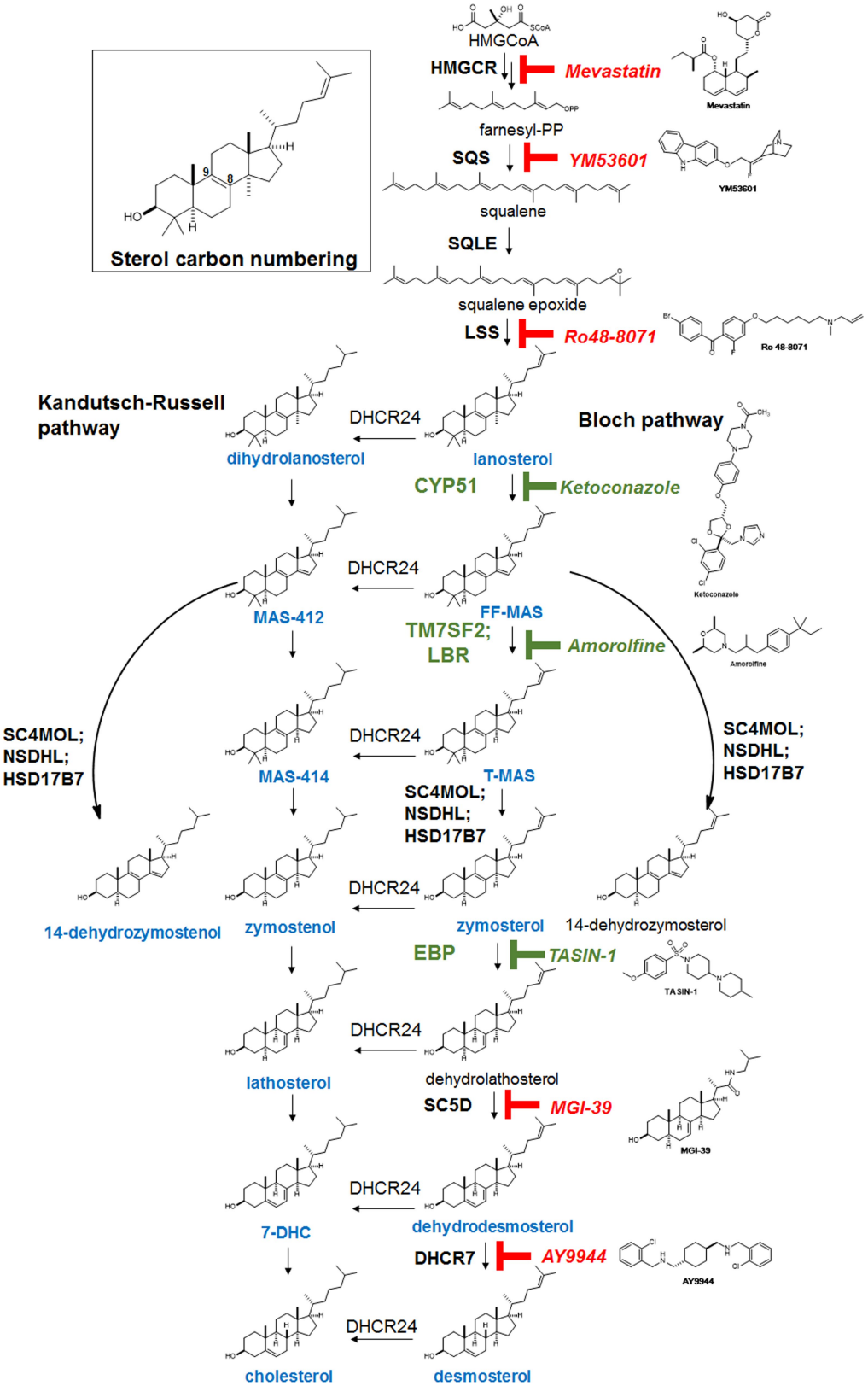
ERα (NR3A1); native receptor; ER GRE-luciferase
ERRα (NR3B1); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
GR (NR3C1); native receptor; GR GRE-luciferase
LXRβ (NR1H2); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
PGR (NR3C3); native receptor; PGR GRE-luciferase
PPARδ (NR1C2); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
PPARγ (NR1C3); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
RARα (NR1B1); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
RARγ (NR1B3); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
RORα (NR1F1); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
RORγ (NR1F3); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
RXRα (NR2B1); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
RXRβ (NR2B2); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
TRα (NR1A1); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
TRβ (NR1A2); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
VDR (NR1I1); Gal4 DBD hybrid receptor; Gal4 UAS-luciferase
NF-κB; native NF-κB; NF-κB GRE-luciferase
NRF2; native receptor; ARE-luciferase

Animal welfare. All animal experiments were performed in accordance with protocols approved by the Case Western Reserve University and George Washington University Institutional Animal Care and Use Committees.

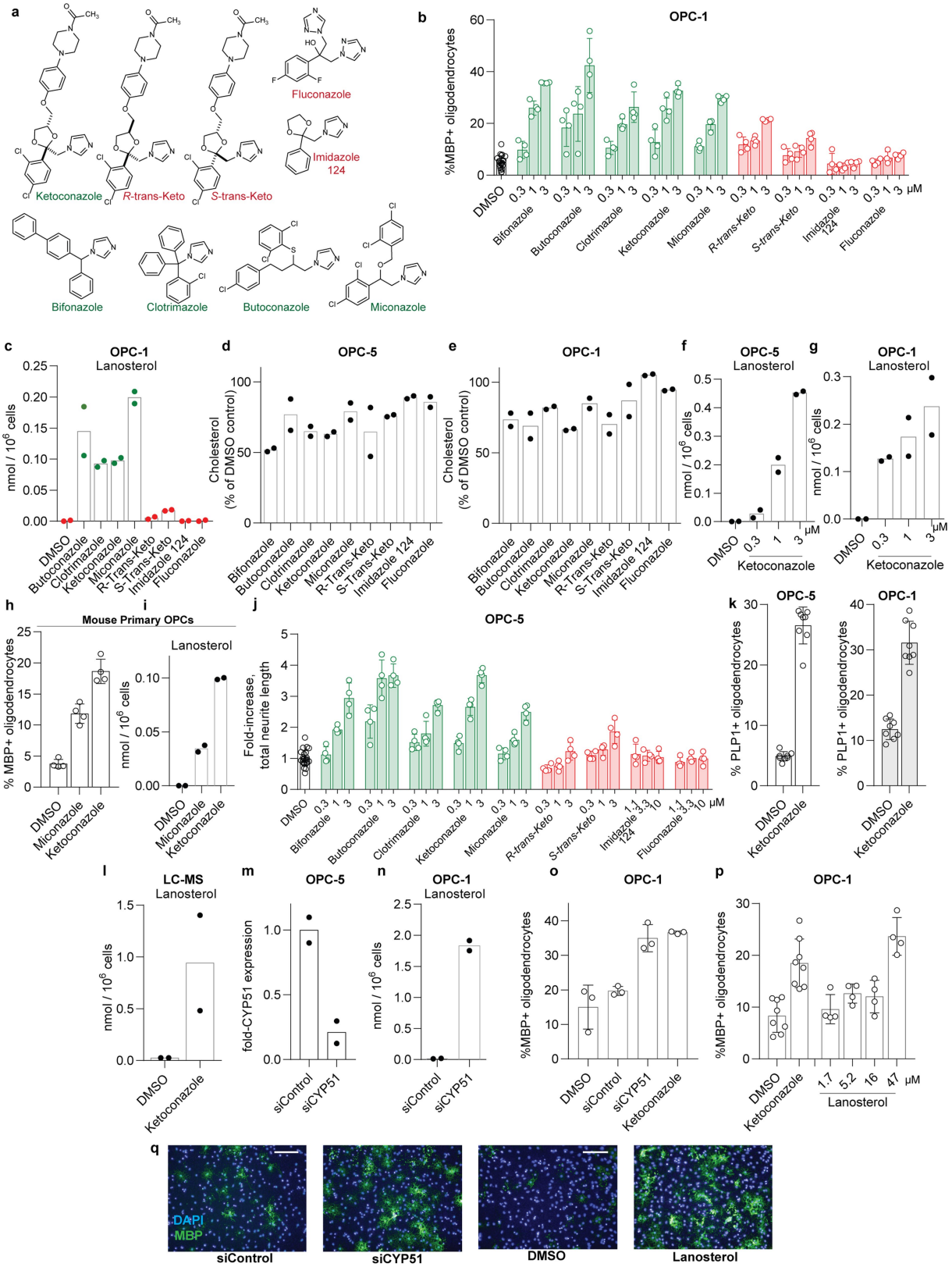
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data supporting the findings of this study are available within the paper (and its Supplementary Information) or from the corresponding author upon request. Source Data for all GC–MS-based sterol profiling experiments and animal experiments are provided with the paper.

31. Godefroi, E. F., Heeres, J., Van Cutsem, J. & Janssen, P. A. The preparation and antimycotic properties of derivatives of 1-phenethylimidazole. *J. Med. Chem.* **12**, 784–791 (1969).
32. Giera, M., Renard, D., Plössl, F. & Bracher, F. Lathosterol side chain amides: a new class of human lathosterol oxidase inhibitors. *Steroids* **73**, 299–308 (2008).
33. Najm, F. J. et al. Rapid and robust generation of functional oligodendrocyte progenitor cells from epiblast stem cells. *Nat. Methods* **8**, 957–962 (2011).
34. Miller, T. E. et al. Transcription elongation factors represent in vivo cancer dependencies in glioblastoma. *Nature* **547**, 355–359 (2017).
35. Honda, A. et al. Highly sensitive quantification of key regulatory oxysterols in biological samples by LC-ESI-MS/MS. *J. Lipid Res.* **50**, 350–357 (2009).
36. Warrilow, A. G., Parker, J. E., Kelly, D. E. & Kelly, S. L. Azole affinity of sterol 14α-demethylase (CYP51) enzymes from *Candida albicans* and *Homo sapiens*. *Antimicrob. Agents Chemother.* **57**, 1352–1360 (2013).
37. Folch, J., Lees, M. & Sloane Stanley, G. H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **226**, 497–509 (1957).
38. Pink, J. J. & Jordan, V. C. Models of estrogen receptor regulation by estrogens and antiestrogens in breast cancer cell lines. *Cancer Res.* **56**, 2321–2330 (1996).
39. Labarca, C. & Paigen, K. A simple, rapid, and sensitive DNA assay procedure. *Anal. Biochem.* **102**, 344–352 (1980).
40. Bosch, D. G. et al. NR2F1 mutations cause optic atrophy with intellectual disability. *Am. J. Hum. Genet.* **94**, 303–309 (2014).

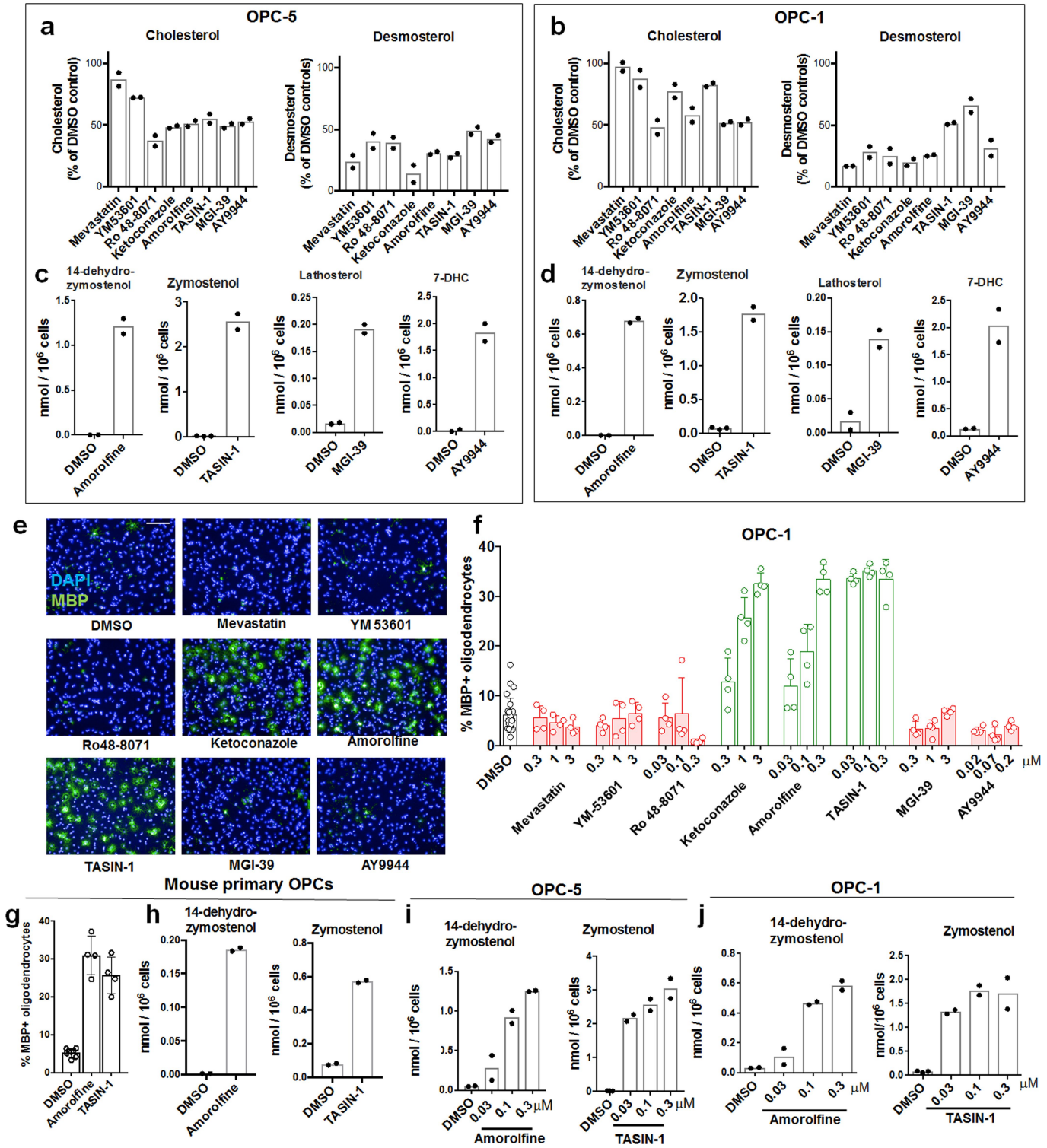


Extended Data Fig. 1 | Expanded cholesterol synthesis pathway diagram. The cascade cyclization of squalene epoxide, catalysed by lanosterol synthase (LSS), provides the first sterol, lanosterol. Processing of lanosterol to cholesterol can proceed via the Kandutsch–Russell and/or Bloch pathways, which use the same enzymes and process substrates that vary only in the presence or absence of the C24 double bond. Intermediates in blue have been confirmed in our GC–MS-based sterol profiling assay using authentic standards. Sterol 14-reductase activity in mouse is shared by two genes, TM7SF2 and LBR. Consistent with past reports[21], inhibition of sterol 14-reductase activity can lead to accumulation of the expected upstream intermediate (FF-MAS) or 14-dehydrozymostenol, also known as cholesta-8,14-dien-3-β-ol. Green indicates enzyme targets and small molecules whose inhibition promotes oligodendrocyte formation.



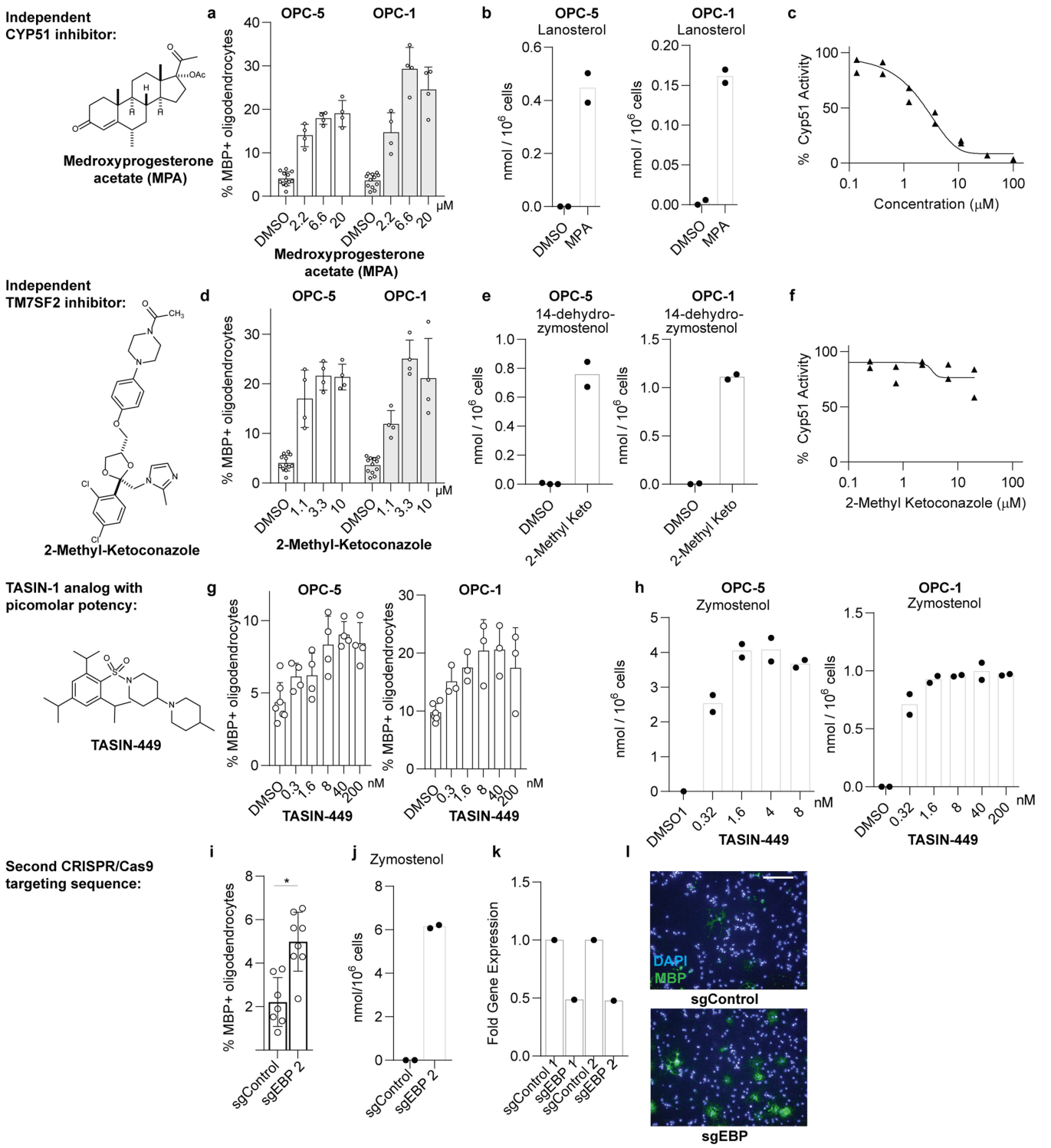
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | CYP51 is the functional target by which imidazole antifungals enhance oligodendrocyte formation. a, Azole molecules with varying degrees of potency for mammalian CYP51 inhibition. Throughout, green labels indicate molecules considered active, while red labels indicate inactive molecules. **b**, Percentage of MBP^+ oligodendrocytes generated from a second, independent derivation of OPCs (OPC-1) at 72 h following treatment with the indicated concentrations of azoles. $n = 4$ wells per condition except DMSO ($n = 24$), with >1,000 cells analysed per well. **c**, GC–MS-based quantification of lanosterol levels in a second derivation of OPCs (OPC-1) treated for 24 h with the indicated azoles at 2.5 μM. $n = 2$ wells per condition. **d**, **e**, GC–MS-based quantification of cholesterol levels in OPCs (OPC-5 and OPC-1) treated for 24 h with the indicated azoles at 2.5 μM. $n = 2$ wells per condition. **f**, **g**, GC–MS-based quantification of lanosterol levels in OPCs (OPC-5, OPC-1) treated for 24 h with the indicated doses of ketoconazole. $n = 2$ wells per condition. Concentrations shown in **f** and **g** mirror those shown in **b** and Fig. 1c. **h**, Percentage of MBP^+ oligodendrocytes generated from mouse primary OPCs at 72 h following treatment with the indicated imidazole antifungals at 3 μM. $n = 4$ wells per condition, with >1,000 cells analysed per well. **i**, GC–MS-based quantification of lanosterol levels in mouse primary OPCs treated for 24 h with the indicated imidazole antifungals at 3 μM. $n = 2$ wells per condition. **j**, Assessment of oligodendrocyte formation using an alternative image quantification metric, fold increase in total neurite length. Re-analysis of data shown in Fig. 1c. $n = 4$ wells per condition except DMSO ($n = 24$), with >1,000 cells analysed per well. **k**, Percentage of oligodendrocytes generated from OPCs at 72 h following treatment with ketoconazole (2.5 μM) as measured by PLP1 immunostaining. Left, OPC-5; right, OPC-1. $n = 8$ wells per condition, with >1,000 cells analysed per well. **l**, LC–MS-based quantification of lanosterol levels in OPC-5 cells treated for 24 h with ketoconazole at 2.5 μM. $n = 2$ wells per condition. **m**, *CYP51* mRNA levels measured by RT–qPCR following 96-h treatment with non-targeting or *CYP51*-targeting pools of cell-permeable siRNAs. $n = 2$ wells per condition. **n**, GC–MS-based quantification of lanosterol levels in OPC-1 cells treated for 96 h with the indicated pooled siRNA reagents. $n = 2$ wells per condition. **o**, Percentage of MBP^+ oligodendrocytes generated from a second, independent batch of OPCs (OPC-1) at 72 h following treatment with the indicated reagents. $n = 3$ wells per condition, with >1,000 cells analysed per well. **p**, Percentage of MBP^+ oligodendrocytes generated from an independent derivation of OPCs at 72 h following treatment with exogenous lanosterol. $n = 4$ wells per condition except DMSO and ketoconazole ($n = 8$), with >1,000 cells analysed per well. **q**, Representative images of OPC-5 cells treated for 72 h with the indicated siRNA reagents and lanosterol. Nuclei are labelled with DAPI (blue), and oligodendrocytes are indicated by immunostaining for MBP (green). Scale bar, 100 μm. All bar graphs indicate mean ± s.d.; **b**, **d**, **h**, **i**, **k**, **l**, **o** and **p** are representative of two independent experiments, and all findings have been confirmed in a second independent derivation of OPCs (Fig. 1).



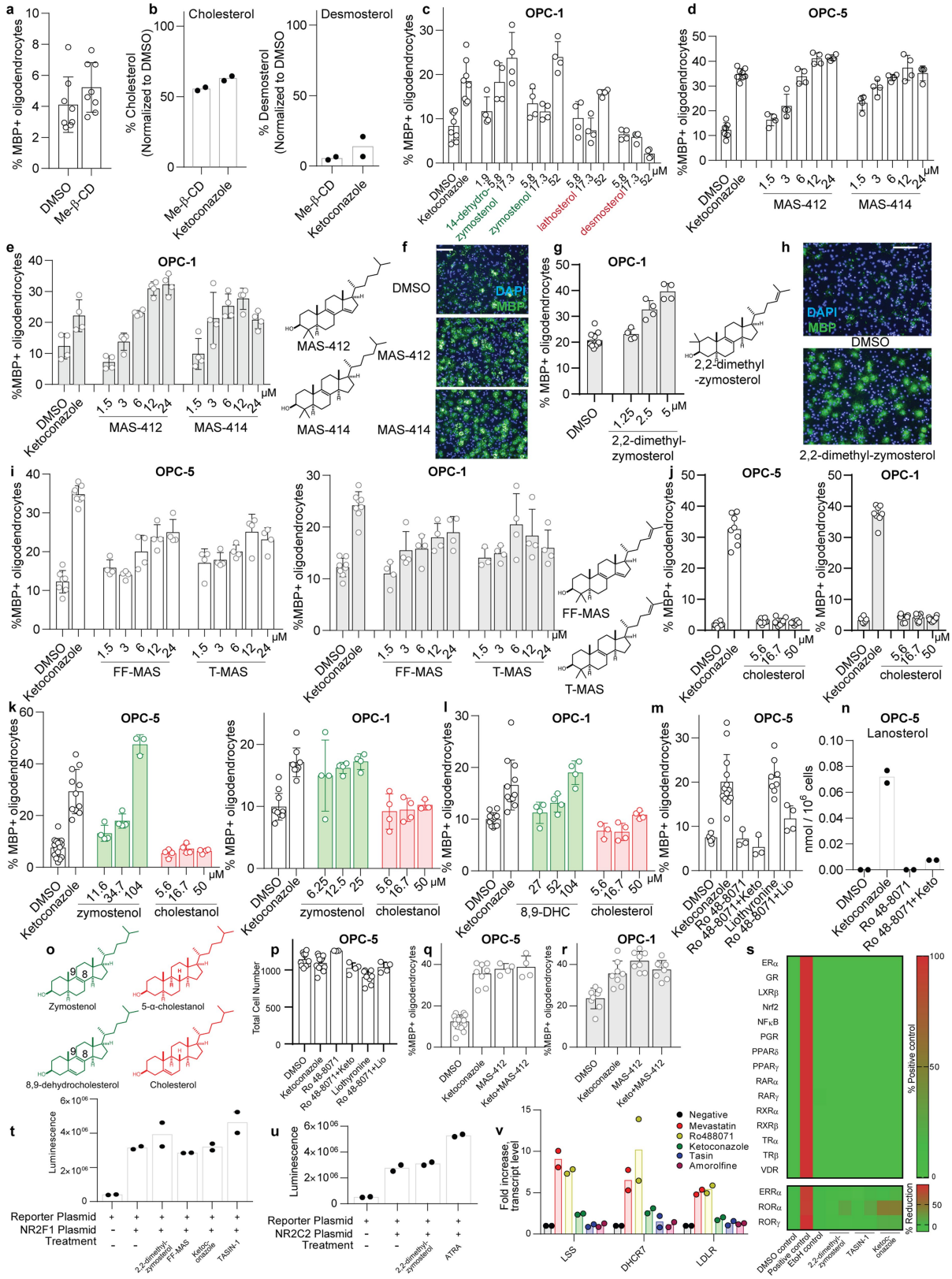
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Effect of small-molecule inhibition of the cholesterol biosynthesis pathway on enhancing oligodendrocyte formation. a, GC–MS-based quantification of sterol levels in OPCs (OPC-5) treated for 24 h with the indicated inhibitors of cholesterol biosynthesis. Left, cholesterol; right, desmosterol. $n = 2$ wells per condition. Inhibitors were used at the following doses unless otherwise noted: mevastatin, ketoconazole, MGI-39, 2.5 μM; YM53601, 2 μM; Ro 48-8071, amorolfine, TASIN-1, 100 nM; AY9944, 200 nM. **b**, GC–MS-based quantification of sterol levels in a second derivation of OPCs (OPC-1). Left, cholesterol; right, desmosterol. $n = 2$ wells per condition. **c**, GC–MS-based quantification of the sterol intermediates expected to accumulate following treatment of OPCs with the indicated inhibitors of cholesterol biosynthesis for 24 h. $n = 2$ wells per condition. **d**, GC–MS-based quantification of the sterol intermediates expected to accumulate following treatment of a second derivation of OPCs (OPC-1) with the indicated inhibitors of cholesterol biosynthesis for 24 h. $n = 2$ wells per condition. In **c** and **d**, no accumulation of other sterol intermediates indicative of off-target effects within the cholesterol pathway were observed (see Source Data). **e**, Representative images of OPC-5 cells treated for 72 h with the indicated small molecules. All treatments are at the highest concentration shown in Fig. 2b. Scale bar, 100 μm. **f**, Percentage of MBP^+ oligodendrocytes generated from a second batch of OPCs (OPC-1) at 72 h following treatment with the indicated cholesterol pathway inhibitors. $n = 4$ wells per condition, except DMSO, $n = 24$, with >1,000 cells analysed per well. **g**, Percentage of MBP^+ oligodendrocytes generated from mouse primary OPCs at 72 h following treatment with the indicated cholesterol pathway inhibitors at 300 nM. $n = 4$ wells per condition, except DMSO, $n = 8$, with >1,000 cells analysed per well. **h**, GC–MS-based quantification of sterol intermediate levels in mouse primary OPCs treated for 24 h with the indicated inhibitors of cholesterol biosynthesis at 300 nM. Left, 14-dehydrozymostenol levels following treatment with amorolfine; right, zymostenol levels following treatment with TASIN-1. $n = 2$ wells per condition. **i**, **j**, GC–MS-based quantification of sterol intermediate levels in OPC-5 (**i**) and OPC-1 (**j**) cells treated for 24 h with the indicated doses of inhibitors of cholesterol biosynthesis. Left, 14-dehydrozymostenol levels following treatment with amorolfine; right, zymostenol levels following treatment with TASIN-1. $n = 2$ wells per condition. Concentrations shown in **i** mirror those shown in **f**. All bar graphs indicate mean ± s.d., and **a**, **c**, **e**–**h** are representative of two independent experiments.



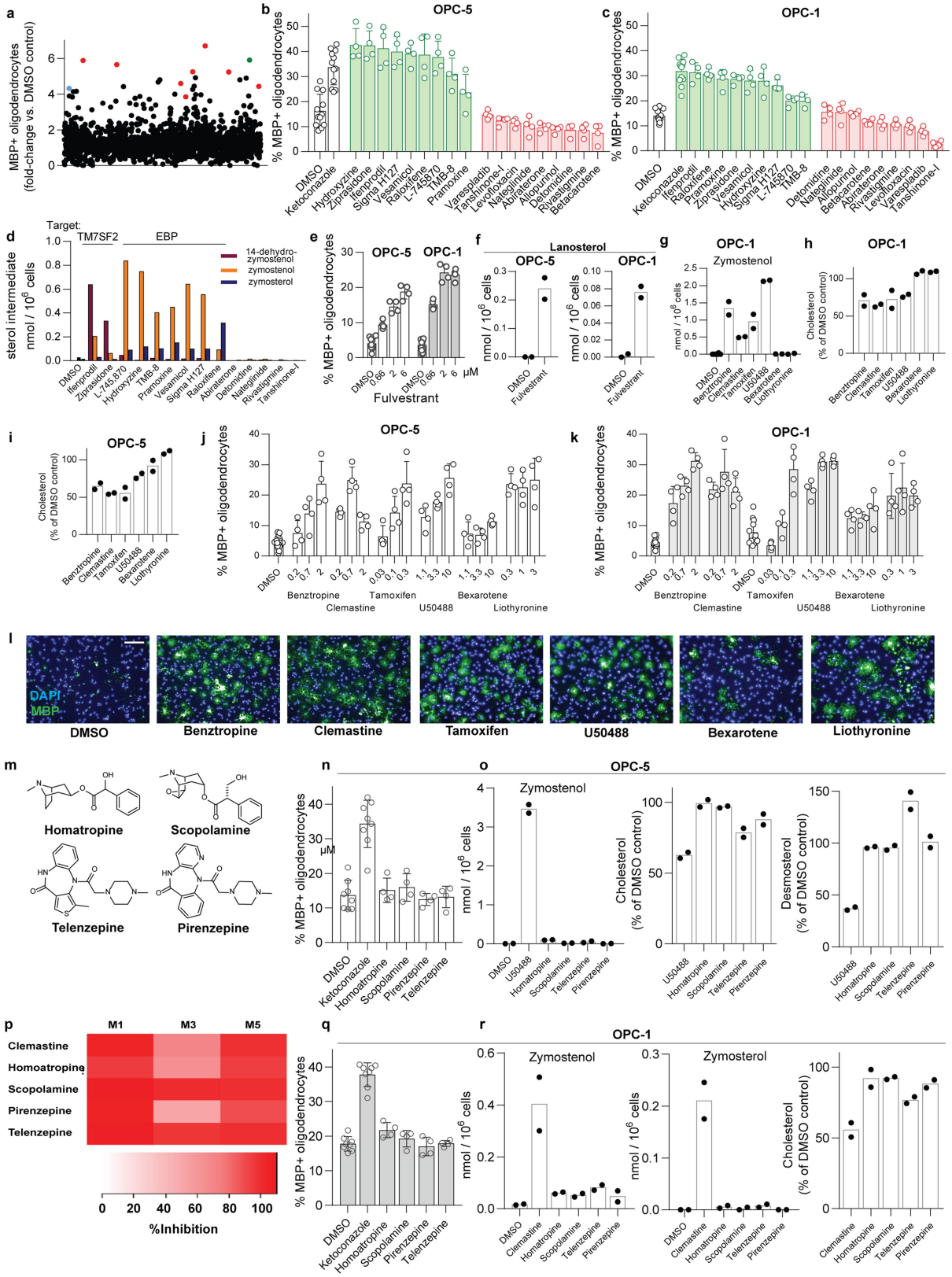
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Effect of independent chemical-genetic and genetic modulators of CYP51, sterol 14 reductase and EBP on oligodendrocyte formation and cholesterol biosynthesis. **a**, **d**, **g**, Percentage of MBP^+ oligodendrocytes generated from two independent derivation of OPCs at 72 h following treatment with the indicated concentrations of medroxyprogesterone acetate (**a**), 2-methyl ketoconazole (**d**) or TASIN-449 (**g**). $n = 4$ wells per condition, except DMSO, $n = 12$ in **a**, **d**. In **g**, for OPC-5, $n = 4$ except DMSO, $n = 7$; for OPC-1, $n = 3$ except DMSO, $n = 6$. **b**, **e**, **h**, GC–MS-based quantification of sterol levels in two independent derivations of OPCs treated for 24 h with medroxyprogesterone acetate at 10 μM (**b**), 2-methyl ketoconazole at 2.5 μM (**e**) and TASIN-449 at the indicated concentrations (**h**). $n = 2$ wells per condition. **c**, **f**, Rat CYP51 enzymatic activity following treatment with varying concentrations of medroxyprogesterone acetate (**c**) and 2-methyl ketoconazole (**f**) as measured by LC–MS-based quantification of the CYP51 product FF-MAS. $n = 2$ independent enzymatic assays. **i**, Percentage of MBP^+ oligodendrocytes generated from OPCs (OPC-5) infected with lentivirus expressing Cas9 and an independent guide RNA targeting EBP (see also Fig. 2c). Eight wells per condition, with >1,000 cells analysed per well. Two-tailed Student's *t*-test, $^*P = 0.0009$. **j**, Functional validation of CRISPR-based targeting of EBP with a second sgRNA using GC–MS-based quantification of zymostenol levels. $n = 2$ wells per condition. **k**, EBP mRNA levels measured by RT–qPCR in OPCs (OPC-5) infected with lentivirus expressing Cas9 and either of two guide RNAs targeting EBP. One well per condition, with results validated in an independent experiment. **l**, Representative images of the oligodendrocyte formation assay shown in Fig. 2c. Nuclei are labelled with DAPI (blue), and oligodendrocytes are indicated by immunostaining for MBP (green). Scale bar, 100 μm. All bar graphs indicate mean ± s.d., and **a**, **d**, **g**, **i**, **k** are representative of two independent experiments.



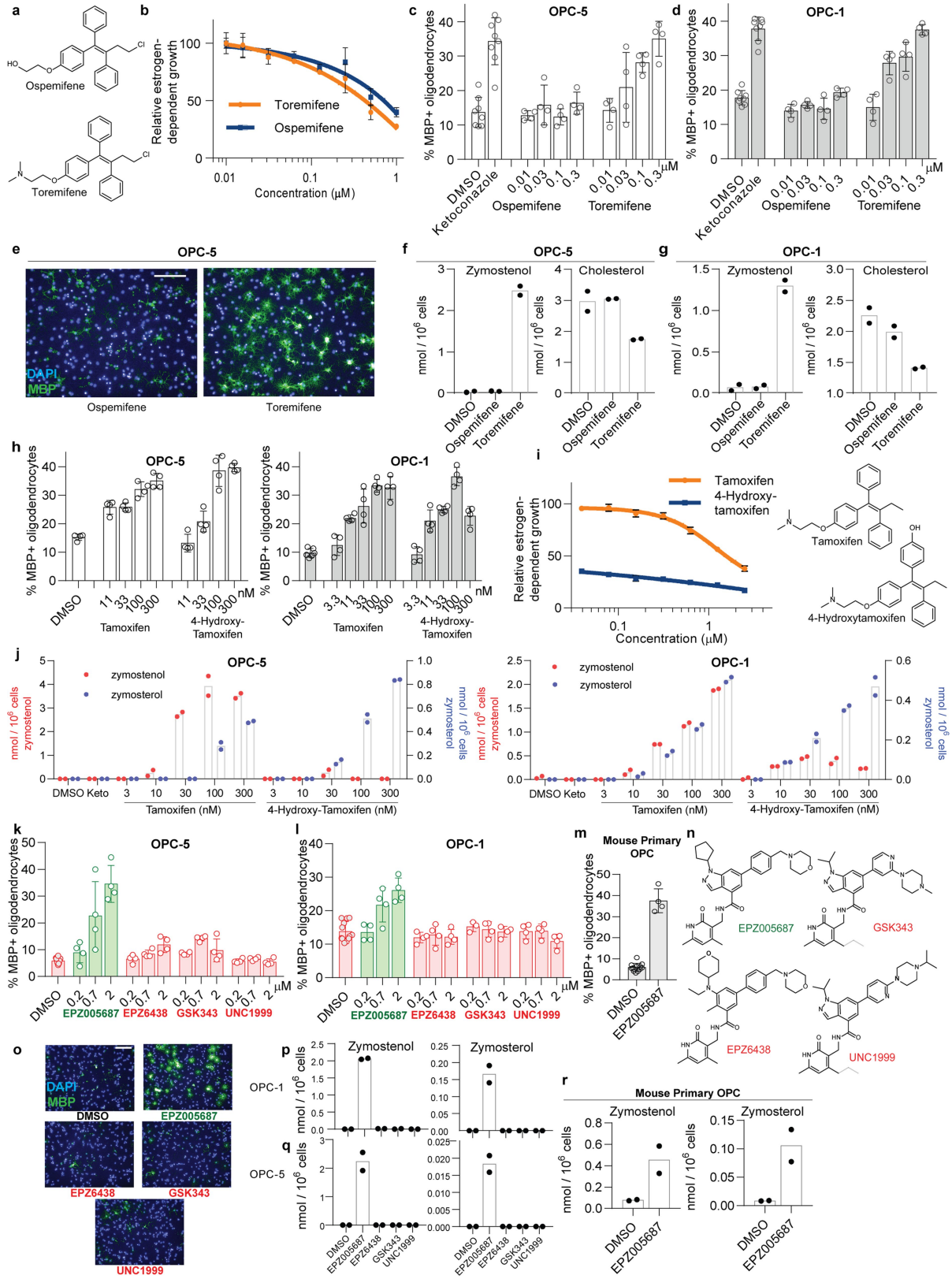
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Effect of 8,9-unsaturated sterols on oligodendrocyte formation. a, Percentage of MBP^+ oligodendrocytes generated from OPCs (OPC-5) at 72 h following treatment with methyl β-cyclodextrin (1 mM) for 30 min at 37 °C. $n = 8$ wells per condition, with >1,000 cells analysed per well. **b**, GC–MS-based quantification of cholesterol (left) and desmosterol (right) in OPCs (OPC-5) treated with methyl β-cyclodextrin (Me-β-CD) at 1 mM or ketoconazole at 2.5 μM. $n = 2$ wells per condition. **c**, **d**, Percentage of MBP^+ oligodendrocytes generated from OPC-1 (**c**) and OPC-5 cells (**d**) at 72 h following treatment with the indicated purified sterol intermediates. $n = 4$ wells per condition, except $n = 8$ for DMSO and ketoconazole, with >1,000 cells analysed per well. Green text highlights metabolites that accumulate after treatments that enhance oligodendrocyte formation (Fig. 2e, Extended Data Fig. 3c). **e**, Percentage of MBP^+ oligodendrocytes generated from OPC1 cells at 72 h following treatment with MAS-412 and MAS-414. $n = 4$ wells per condition, with >1,000 cells analysed per well. **f**, Representative images of OPC5 cells treated for 72 h with DMSO, MAS-412, or MAS-414 (3 μM). Nuclei are labelled with DAPI (blue), and oligodendrocytes are indicated by immunostaining for MBP (green). Scale bar, 100 μm. **g**, Percentage of MBP^+ oligodendrocytes generated from OPC-1 at 72 h following treatment with 2,2-dimethyl-zymosterol. $n = 4$ wells per condition except DMSO ($n = 12$), with >1,000 cells analysed per well. **h**, Representative images of OPC-5 cells treated for 72 h with vehicle and 2,2-dimethyl-zymosterol (2.5 μM). Nuclei are labelled with DAPI (blue), and oligodendrocytes are indicated by immunostaining for MBP (green). Scale bar, 100 μm. **i**, Percentage of MBP^+ oligodendrocytes generated from OPC-5 (left) and OPC-1 (right) cells at 72 h following treatment with FF-MAS or T-MAS. $n = 4$ wells per condition except DMSO and ketoconazole ($n = 8$), with >1,000 cells analysed per well. **j**, Percentage of MBP^+ oligodendrocytes generated from OPC-5 and OPC-1 OPCs at 72 h following treatment with the indicated concentrations of cholesterol. $n = 8$ wells per condition, with >1,000 cells analysed per well. **k**, **l**, Percentage of MBP^+ oligodendrocytes generated from OPC-5 and OPC-1 cells at 72 h following treatment with the indicated concentrations of sterols that are structurally identical aside from the presence or absence of the 8,9 double bond (structures in **o**). $n \geq 3$ wells per condition (see dot plots as replicate values vary by condition), with >1,000 cells analysed per well. **m**, Percentage of MBP^+ oligodendrocytes generated from OPCs (OPC-5) at 72 h following treatment with the indicated small molecules or combinations of small molecules (ketoconazole, 2.5 μM; Ro 48-8071, 11 nM; liothyronine, 3 μM). $n = 3$ wells per condition, except DMSO $n = 11$, ketoconazole $n = 13$, liothyronine $n = 8$ & liothyronine + Ro 48-8071 $n = 4$, with >1,000 cells analysed per well. **n**, GC–MS-based quantification of lanosterol levels in OPCs (OPC-5) treated for 24 h with the indicated small molecules or combinations of small molecules at concentrations stated in **m**. $n = 2$ wells per condition. **o**, Structures of zymostenol, 8,9-dehydrocholesterol, 5α-cholestanol, and cholesterol. **p**, Total cell number as measured by counting of $DAPI^+$ nuclei in the experiment presented in **m**. **q**, **r**, Percentage of MBP^+ oligodendrocytes generated from OPCs (OPC5 and OPC-1) at 72 h following treatment with the indicated small molecules or combinations of small molecules in two independent batches of OPCs (ketoconazole, 2.5 μM; MAS412, 5 μM). In **q**, $n = 16$ for DMSO, 8 for ketoconazole, and 4 for remaining bars. In **r**, $n = 8$ wells per condition. **s**, Luciferase reporter assays were used to assess whether 2,2-dimethylzymosterol (5 μM), ketoconazole (2.5 μM), and TASIN-1 (250 nM) modulate human ERα, GR, LXRβ, NFkB, NRF2, PGR, PPARδ, PPARγ, RARα, RARγ, RXRα, RXRβ, TRα, TRβ and VDR transcriptional activity in agonist mode and ERRα, RORα and RORγ in inverse-agonist mode. $n = 2$ wells per condition and $n = 3$ wells per positive control condition. **t**, Effects of sterols (2,2-dimethylzymosterol 5 μM, FF-MAS 10 μM) and small molecules (ketoconazole 2.5 μM, TASIN-1 100 nM) on the NR2F1-mediated activation of a NGFI-A promoter driven luciferase reporter. $n = 2$ wells per condition. **u**, Effects of 2,2-dimethylzymosterol (5 μM) on NR2C2-mediated activation of a NGFI-A promoter driven luciferase reporter in comparison to cells transfected with reporter only, untreated, or treated with a previously reported positive control (all-*trans* retinoic acid, ATRA, 5 μM). $n = 2$ wells per condition. **v**, LSS, DHCR7, LDLR mRNA levels measured by RT–qPCR following 24 h treatment with DMSO, mevastatin (2.5 μM), Ro 48-8071 (500 nM), ketoconazole (2.5 μM), TASIN-1 (100 nM), or amorolfine (100 nM). $n = 2$ wells. All bar graphs indicate mean ± s.d., and **a**–**n** and **t**–**v** are representative of two independent experiments.



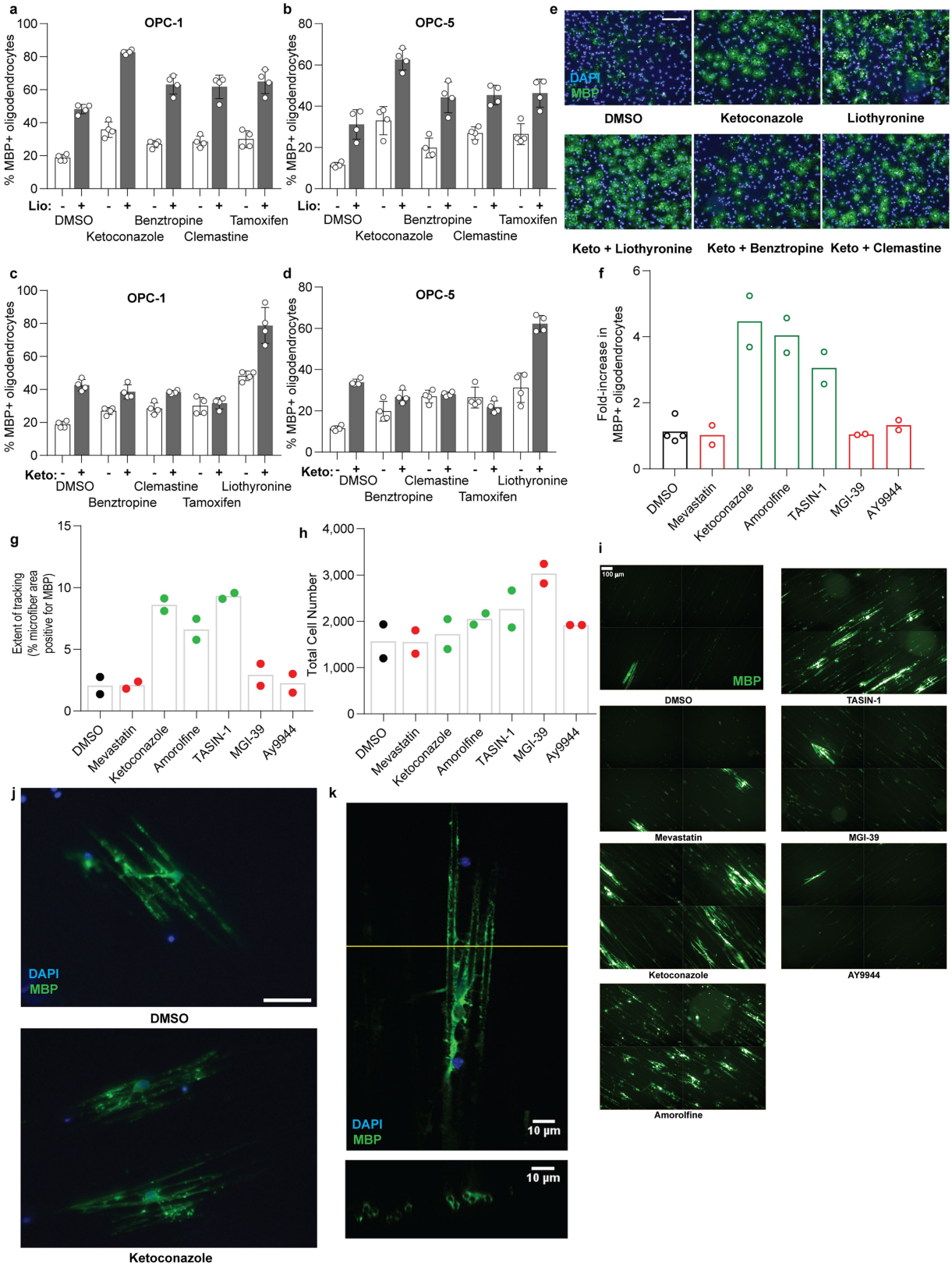
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Inhibiting CYP51, TM7SF2 and EBP is a unifying mechanism for many small-molecule enhancers of oligodendrocyte formation identified by high-throughput screening. **a**, Percentage of MBP^+ oligodendrocytes (relative to DMSO control wells) generated from OPCs (OPC-1 derivation) at 72 h following treatment with a library of 3,000 bioactive small molecules, each at 2 μM. Each dot represents the result for one small molecule in the library. Red, imidazole antifungals; blue, clemastine; green, EPZ005687, the top novel hit molecule (Extended Data Fig. 7). **b**, **c**, Percentage of MBP^+ oligodendrocytes generated from OPCs (left: OPC-5; right: OPC-1) at 72 h following treatment with ketoconazole, nine top molecules identified by bioactives screening (green), and nine randomly chosen library members (red) at a uniform dose of 5 μM. $n = 4$ wells per condition except DMSO and ketoconazole, $n = 12$ wells, with >1,000 cells analysed per well. **d**, GC–MS-based quantification of zymosterol, zymostenol, and 14-dehydrozymostenol levels in a second batch of OPCs treated for 24 h with the indicated screening hits and randomly chosen library members at 2 μM. $n = 1$; for validation in a second derivation of OPCs, see Fig. 3a. Molecules are clustered by enzyme targeted (top labels). **e**, Percentage of MBP^+ oligodendrocytes generated from OPCs at 72 h following treatment with the indicated doses of fulvestrant, one of the top 10 HTS hits. $n = 4$ wells per condition except DMSO, $n = 12$, with >1,000 cells analysed per well. **f**, GC–MS-based quantification of lanosterol levels in OPCs treated for 24 h with fulvestrant at 2 μM. $n = 2$ wells per condition. **g**–**i**, GC–MS-based quantification of metabolite levels in OPCs treated for 24 h with the indicated previously reported enhancers of oligodendrocyte formation at the following doses: benztropine, 2 μM; clemastine, 1 μM; tamoxifen, 100 nM; U50488, 5 μM; bexarotene, 1 μM; liothyronine, 3 μM. $n = 2$ wells per condition. **j**, **k**, Percentage of MBP^+ oligodendrocytes generated from OPCs (OPC-5 left, OPC-1 right) at 72 h following treatment with the indicated previously reported enhancers of oligodendrocyte formation. $n = 4$ wells per condition, except DMSO $n = 20$ for OPC-5 and $n = 12$ for OPC-1, with >1,000 cells analysed per well. All doses are in μM. **l**, Representative images of OPCs treated for 72 h with the indicated small molecules. All treatments in **l** are at the highest concentration shown in **j**. Scale bar, 100 μm. **m**, Structures of muscarinic receptor antagonists used in this study. **n**, **q**, Percentage of MBP^+ oligodendrocytes generated from OPCs (OPC-5: top, OPC-1: bottom) at 72 h following treatment with ketoconazole or the indicated muscarinic receptor modulators at 2 μM, the concentration used during screening. $n = 4$ wells per condition except DMSO and ketoconazole, $n = 8$, with >1,000 cells analysed per well. **o**, GC–MS-based quantification of three metabolite levels in OPC-5 OPCs treated for 24 h with U50488 (5 μM) or the indicated muscarinic receptor modulators (2 μM). Left, zymostenol; centre, cholesterol; right, desmosterol. $n = 2$ wells per condition. **p**, Heatmap indicating inhibition of muscarinic receptor isoforms M1, M3, and M5 by the indicated small molecules (2 μM) assayed using GeneBLAzer NFAT-*bla* CHO-K1 cells. $n = 2$ wells per condition. **r**, GC–MS-based qua*n*tification of three metabolite levels in OPC-1 OPCs treated for 24 h with clemastine (1 μM) or the indicated muscarinic receptor modulators at 2 μM. $n = 2$ wells per condition. Left, zymostenol; centre, zymosterol; right, cholesterol. Sigma H127, *p*-fluorohexahydro-sila-difenidol. All bar graphs indicate mean ± s.d., and **b**, **c**, **e**, **i**, **j**, **k**, **n**, **q** are representative of two independent experiments.



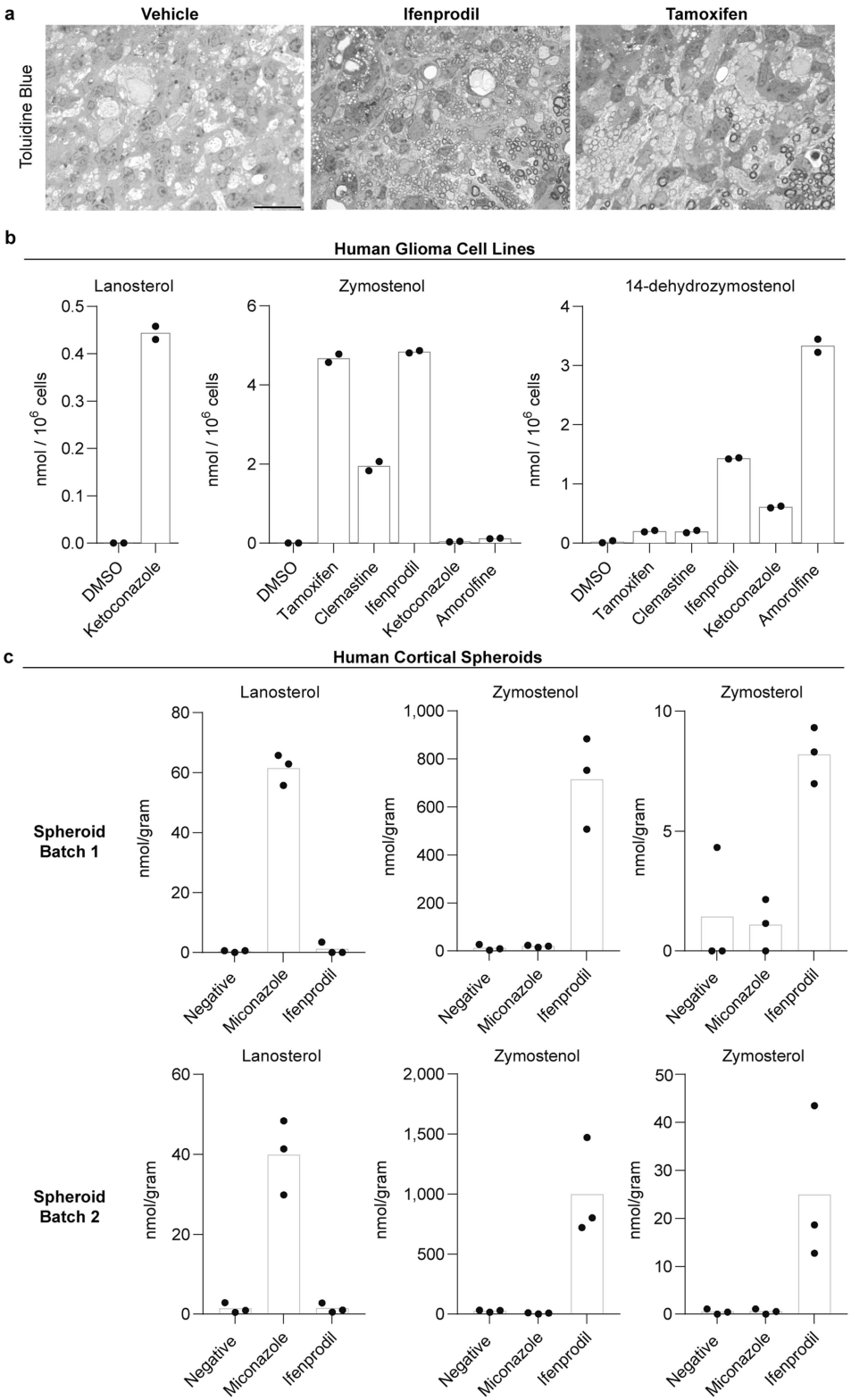
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Effect of selective oestrogen receptor modulators and EZH2 inhibitors on cellular EBP function and oligodendrocyte formation. a, Structures of selective oestrogen receptor modulators used in this study. **b**, Effects of ospemifene and toremifene on the oestrogen-dependent growth of T47D cells. $n = 3$ wells per condition. **c**, **d**, Percentage of MBP^+ oligodendrocytes generated from two independent batches of OPCs at 72 h following treatment with ospemifene and toremifene. $n = 4$ wells per condition except DMSO and ketoconazole, $n = 8$, with >1,000 cells analysed per well. **e**, Representative images of OPCs treated for 72 h with the indicated small molecules. All molecules were used at 300 nM. Scale bar, 100 μm. **f**, **g**, GC–MS-based quantification of two metabolite levels in OPCs treated for 24 h with ospemifene and toremifene at 300 nM. Left, zymostenol; right, cholesterol. $n = 2$ wells per condition. **h**, Percentage of MBP^+ oligodendrocytes generated from two independent batches of OPCs at 72 h following treatment with tamoxifen and 4-hydroxytamoxifen. Left, OPC-5; right, OPC-1. $n = 4$ wells per condition, except DMSO, $n = 6$ for OPC-1 (right). **i**, Effects of tamoxifen and 4-hydroxytamoxifen on the oestrogen-dependent growth of T47D cells. $n = 3$ wells per condition. **j**, GC–MS-based quantification of zymostenol (left axis) and zymosterol levels (right axis) in OPC-5 and OPC-1 treated 24 h with tamoxifen and 4-hydroxytamoxifen at the indicated concentrations. $n = 2$ wells per condition. **k**, Percentage of MBP^+ oligodendrocytes generated from OPCs at 72 h following treatment with the indicated structurally analogous EZH2 inhibitors. $n = 4$ wells per condition, except DMSO, $n = 12$, with >1,000 cells analysed per well. **l**, Percentage of MBP^+ oligodendrocytes generated from a second batch of OPCs at 72 h following treatment with the indicated structurally analogous EZH2 inhibitors. $n = 4$ wells per condition, except DMSO, $n = 12$, with >1,000 cells analysed per well. **m**, Percentage of MBP^+ oligodendrocytes generated from mouse primary OPCs at 72 h following treatment with EPZ005687. $n = 4$ wells per condition, except DMSO, $n = 12$, with >1,000 cells analysed per well. **n**, Structure of EPZ005687 and structurally analogous EZH2 inhibitors. **o**, Representative images of OPCs treated for 72 h with the indicated EZH2 inhibitors. All treatments are at 2 μM. Scale bar, 100 μm. **p**, GC–MS-based quantification of two sterol intermediates following treatment of OPCs with the indicated EZH2 inhibitors at 1 μM for 24 h. Left, zymostenol; right, zymosterol. $n = 2$ wells per condition. **q**, GC–MS-based quantification of two sterol intermediates following treatment of a second derivation of OPCs with the indicated EZH2 inhibitors at 1 μM for 24 h. Left, zymostenol; right, zymosterol. $n = 2$ wells per condition. **r**, GC–MS-based quantification of two sterol intermediates following treatment of mouse primary OPCs with EPZ005687 at 2 μM for 24 h. Left, zymostenol; right, zymosterol. $n = 2$ wells per condition. All bar graphs indicate mean ± s.d., and **c**, **d**, **h**, **k**–**o**, **r** are representative of two independent experiments.

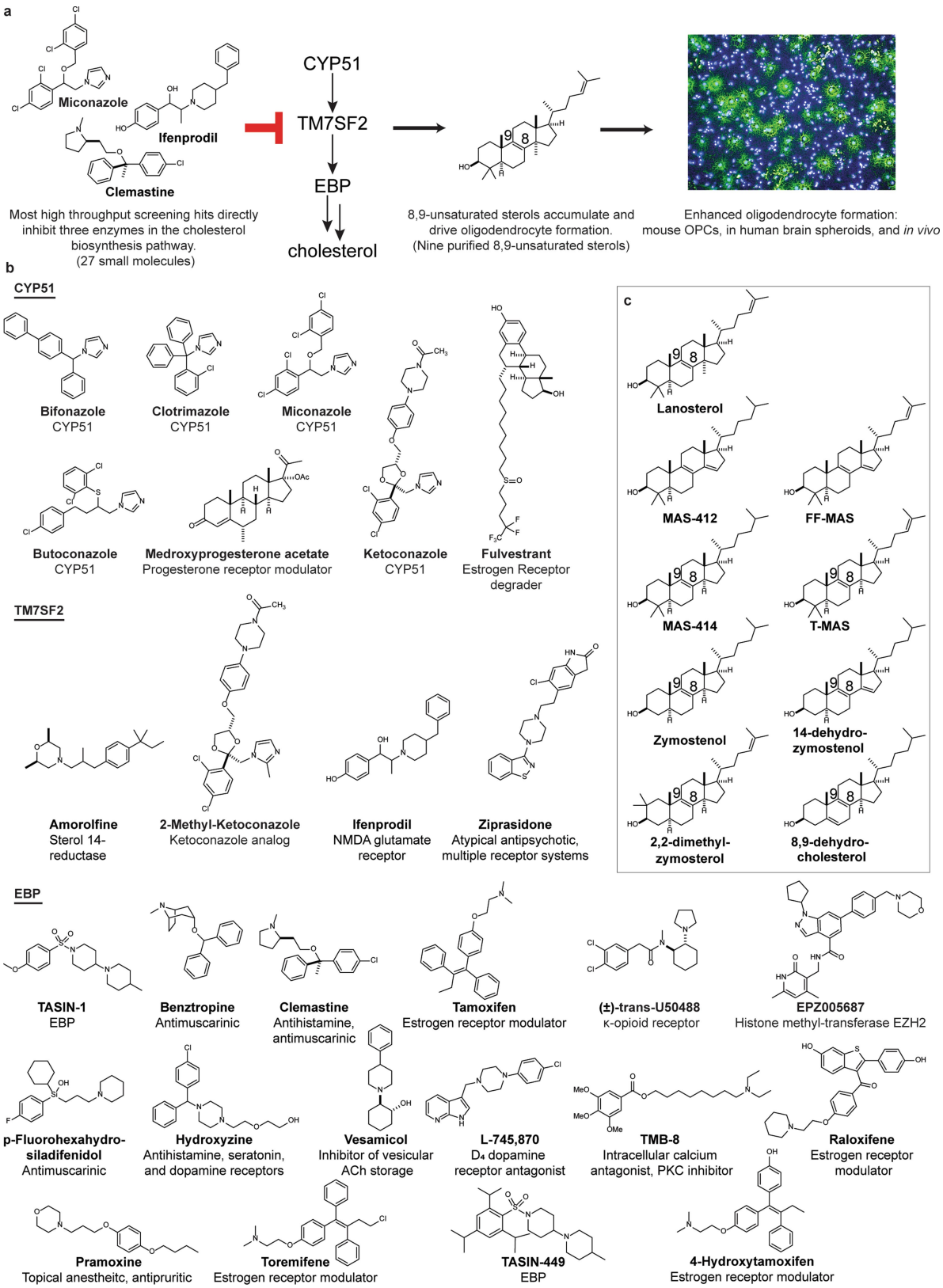


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Effect of combinations of small-molecule treatments on oligodendrocyte formation, and ability of oligodendrocytes to track along and wrap electrospun microfibres after single small-molecule treatments. a, **b**, Percentage of MBP^+ oligodendrocytes generated from OPCs (left, OPC-1; right, OPC-5) at 72 h following treatment with the indicated combinations of liothyronine and enhancers of oligodendrocyte formation. Unless noted, the following concentrations were used: ketoconazole, 2.5 μM; benztropine, 2 μM; clemastine 2 μM; tamoxifen 200 nM; liothyronine, 3 μM. $n = 4$ wells per treatment condition, with >1,000 cells analysed per well. Lio, liothyronine. **c**, **d**, Percentage of MBP^+ oligodendrocytes generated from OPCs at 72 h following treatment with the indicated combinations of ketoconazole and enhancers of oligodendrocyte formation. $n = 4$ wells per treatment condition, with >1,000 cells analysed per well. **e**, Representative images of OPCs treated for 72 h with the indicated small molecules. Small-molecule concentrations are as in a. Scale bar, 100 μm. **f**, Fold-increase in MBP^+ oligodendrocytes following plating of OPCs (OPC-5) onto microfibres and treatment for 14 days with the indicated pathway modulators. $n = 2$ wells per condition, except DMSO, $n = 4$. **g**, In an independent experiment, OPCs (OPC-5) were plated onto microfibres, treated with small molecules for 4 days, and fixed and stained after 14 days. The extent to which MBP^+ oligodendrocytes tracked along the microfibre substrate was measured. $n = 2$ wells per condition. **h**, Total $DAPI^+$ cell number in the experiment in **g**. **i**, Representative images highlighting tracking along the microfibre substrate. Each image is a montage of four separate images within the same well. Green, MBP. Scale bar, 100 μm. **j**, High-resolution images of MBP^+ oligodendrocytes tracking along microfibres. Green, MBP; blue, DAPI. Ketoconazole, 2.5 μM. Scale bar, 50 μm. **k**, Confocal imaging of OPCs seeded onto aligned microfibres and treated for 14 days with ketoconazole (2.5 μM). The plane of the cross-section is highlighted in yellow and the cross-section, in which green fluorescence appears to encircle several microfibres, is shown in the bottom panel. Green, MBP; blue, DAPI. All bar graphs indicate mean ± s.d., and **a**–**d** are representative of two independent experiments.



Extended Data Fig. 9 | Effect of oligodendrocyte-enhancing small molecules on sterol levels in human cells and human cortical spheroids. **a**, Representative images of toluidine blue-stained sections of LPC-lesioned dorsal spinal cord from mice treated for 8 days with ifenprodil (10 mg per kg) or tamoxifen (2 mg per kg). Scale bar, 20 μm. **b**, GC–MS-based quantification of three metabolite levels in human glioma cells (GBM528) treated for 24 h with the indicated small molecules at the following concentrations: tamoxifen, 100 nM; clemastine, 2 μM; ifenprodil, 2 μM; ketoconazole, 2.5 μM; amorolfine, 100 nM. Left, lanosterol; centre, zymostenol; right, 14-dehydrozymostenol. $n = 2$ wells per condition. **c**, GC–MS-based quantification of three metabolite levels in two independent batches of human cortical spheroids treated for 24 h with the indicated small molecules at 2 μM. Left, lanosterol; centre, zymostenol; right, zymosterol. $n = 3$ spheroids per condition; representative of two independent experiments.



Extended Data Fig. 10 | Twenty-seven small molecules and nine purified 8,9-unsaturated sterols shown to enhance the formation of oligodendrocytes. a, Schematic showing the proposed mechanism of action for enhanced oligodendrocyte formation by diverse small molecules. **b**, Molecules that enhance oligodendrocyte formation are grouped by enzyme inhibited (GC–MS analysis in OPCs): CYP51, top; sterol 14-reductase, centre; EBP, bottom. **c**, Purified 8,9-unsaturated sterols that enhance oligodendrocyte formation.

LETTER

https://doi.org/10.1038/s41586-018-0394-6

A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte

Lindsey W. Plasschaert[1,5,7], Rapolas Žilionis[2,3,7], Rayman Choo-Wing[1,5], Virginia Savova[2,6], Judith Knehr[4], Guglielmo Roma[4], Allon M. Klein[2]* & Aron B. Jaffe[1,5]*

The functions of epithelial tissues are dictated by the types, abundance and distribution of the differentiated cells they contain. Attempts to restore tissue function after damage require knowledge of how physiological tasks are distributed among cell types, and how cell states vary between homeostasis, injury–repair and disease. In the conducting airway, a heterogeneous basal cell population gives rise to specialized luminal cells that perform mucociliary clearance[1]. Here we perform single-cell profiling of human bronchial epithelial cells and mouse tracheal epithelial cells to obtain a comprehensive census of cell types in the conducting airway and their behaviour in homeostasis and regeneration. Our analysis reveals cell states that represent known and novel cell populations, delineates their heterogeneity and identifies distinct differentiation trajectories during homeostasis and tissue repair. Finally, we identified a novel, rare cell type that we call the 'pulmonary ionocyte', which co-expresses *FOXI1*, multiple subunits of the vacuolar-type H^+-ATPase (V-ATPase) and *CFTR*, the gene that is mutated in cystic fibrosis. Using immunofluorescence, modulation of signalling pathways and electrophysiology, we show that Notch signalling is necessary and *FOXI1* expression is sufficient to drive the production of the pulmonary ionocyte, and that the pulmonary ionocyte is a major source of CFTR activity in the conducting airway epithelium.

The conducting airway is lined with a pseudostratified epithelium consisting of basal, secretory and ciliated cells, as well as rare pulmonary neuroendocrine cells (PNECs) and brush cells[2]. Studies of lineage tracing and regeneration post-injury show that basal cells are a heterogeneous population that contains the epithelial stem cells[3,4]. Basal cells differ in their expression of cytokeratins 14 and 8 (Krt14 and Krt8) and luminal cell fate determinants that are upregulated upon injury[2,5]. To identify the full repertoire of basal cell molecular states, and to identify candidate gene expression programs that might bias basal cells to self-renew or to adopt differentiated fates, we performed single-cell RNA profiling on airway epithelial cells. We also sought to elucidate the molecular composition of rare PNECs and brush cells, which have fewer lineage markers and are more difficult to define functionally[6,7]. Because our approach is unbiased and comprehensive, it could also identify new cell types with a role in mucociliary clearance.

We performed single-cell RNA sequencing analysis[8] (scRNA-seq) on 7,662 mouse tracheal epithelial cells and 2,970 primary human bronchial epithelial cells (HBECs) differentiated at an air–liquid interface (ALI)[9] (Fig. 1a, b). As there are well-documented differences between mouse and human airways[10], using these two systems enables comparative analyses and prioritization of common findings between mouse and human. This also provided in vivo validation of findings in the culture model, which lacks non-epithelial cells and uses defined culture conditions. A similar analysis of mouse tracheal epithelial cells in the accompanying Article[11] corroborates many of our findings.

We visualized the single-cell data using a graph-based algorithm (SPRING[12]) that conserves neighbouring relationships of gene expression, facilitating analysis of differentiation trajectories. The resulting graphs revealed a non-uniform continuum structure spanning basal-to-luminal differentiation, with rare gene expression states representing satellite clusters (see 'Data availability' in Methods). Using spectral clustering, we partitioned cells into populations with specific, reproducible gene expression signatures (Fig. 1c, d). On the basis of enrichment of previously annotated markers (Supplementary Tables 1, 2), we identified clusters in mouse (Fig. 1c) and human (Fig. 1d) that represented known cell types[2,7]: basal, secretory, ciliated, brush and PNECs. We performed pairwise correlation analysis as a measure of relatedness between clusters, and curated a list of transcription factors, surface molecules and kinases expressed in each cluster (Extended Data Fig. 1, Supplementary Tables 1–3). Our analysis confirmed previous findings[5,13–15] that basal and secretory populations are heterogeneous, and uncovered additional molecular heterogeneity (Extended Data Figs. 2, 3). Basal cells formed a continuum of states defined by gene modules associated with a basal-to-luminal gene expression axis (*Krt5* versus *Krt8*) as well as by variable expression of genes associated with basement membrane deposition and remodelling. In both mouse and human, *Col17a1*/*COL17A1* (gene homologues are written as mouse/human throughout) and IGFBP family members (*Igfbp3*/*IGFBP6*) correlated with the basal cell sub-population marker *Krt14*, whereas in mouse an independent *Krt14*$^-$ module associated with the basal cell adhesion molecule *Bcam* and with *Dcn*—which encodes decorin, a regulator of collagen fibrillogenesis. Among secretory cells, many differences were associated with different levels of maturity, with the least-mature cells expressing basal cell transcripts (for example, *Krt5*/*KRT5* and *Trp63*/*TP63*), and the most-mature cells expressing *Muc5b*/*MUC5B*. Secretory cells also differed in other ways. In human, one cluster associated with antigen presentation (human leukocyte antigen (HLA) gene family members). In mouse, the secretory cells appeared to associate with two distinct trajectories from the basal layer: those expressing *Krt4*, and those emerging from a *Krt4*^{low} state marked by *Trp63, Bcam* and *Dcn*. Heterogeneity of both basal and secretory cells was also associated with tens of other genes with diverse functions, including signalling molecules (for example, *Wnt10a*) and early specifiers of mature lineages (for example, *Foxj1*) (Extended Data Figs. 2, 3).

Our analysis also revealed gene signatures of epithelial cell states that have not previously been described. First, the paired cytokeratins 4 and 13 (*Krt4*/*Krt13*) defined a unique cluster in the mouse dataset located between basal and secretory, suggesting that this may be a transitional cell state (Fig. 1c, Supplementary Table 1). Immunofluorescence of Krt4 in mouse tracheal epithelium demonstrated that it was co-enriched in subsets of Krt5$^+$ basal cells, Krt8$^+$ luminal cells and Scgb1a1$^+$ club (secretory) cells, but not in Foxj1$^+$ ciliated cells (Extended Data Fig. 4a). This pattern is reminiscent of the proposed model for basal luminal precursors (BLPs), a subset of non-transit amplifying basal cells with upregulated luminal markers[14]. In addition, *KRT4*/*KRT13* expression

[1]Chemical Biology & Therapeutics, Novartis Institutes for BioMedical Research, Cambridge, MA, USA. [2]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [3]Institute of Biotechnology, Vilnius University, Vilnius, Lithuania. [4]Chemical Biology & Therapeutics, Novartis Institutes for BioMedical Research, Basel, Switzerland. [5]Present address: Respiratory Diseases, Novartis Institutes for BioMedical Research, Cambridge, MA, USA. [6]Present address: Precision Immunology, Immunology & Inflammation Research Therapeutic Area, Sanofi, Cambridge, MA, USA. [7]These authors contributed equally: Lindsey W. Plasschaert, Rapolas Žilionis. *e-mail: allon_klein@hms.harvard.edu; aron.jaffe@novartis.com

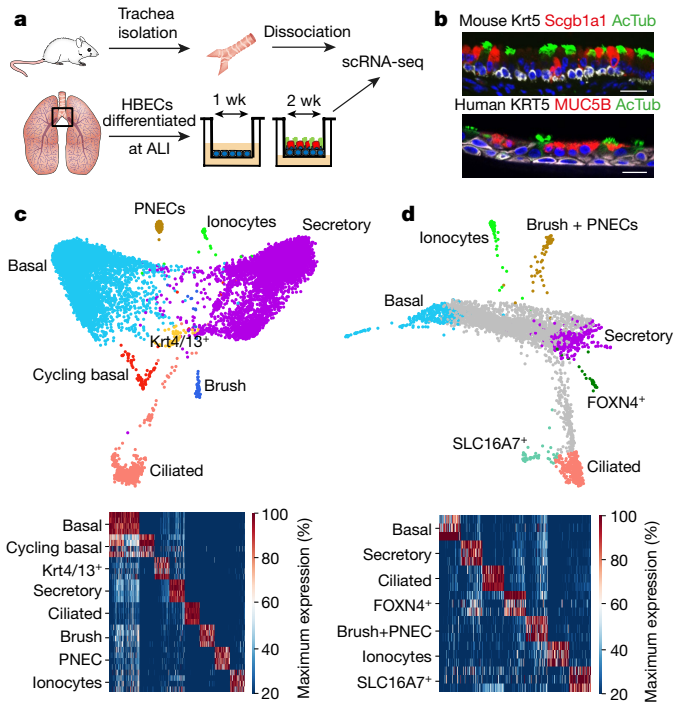


Fig. 1 | Single-cell RNA-seq of proximal airway epithelial cells in mouse and human. a, Mouse tracheal epithelial cells were isolated, dissociated and collected for scRNA-seq. Human bronchial epithelial cells (HBECs) were cultured for 1 week submerged, followed by 2 weeks at an air–liquid interface (ALI) and collected for scRNA-seq. **b**, Mouse tracheal epithelium ($n = 3$ mice) and differentiated HBEC culture ($n = 3$ donors) are pseudostratified, containing basal cells (KRT5), secretory cells (Scgb1a1 in mouse; MUC5B in human), and ciliated cells (AcTub, acetylated α-tubulin). Scale bars, 20 μm. **c**, **d**, SPRING plots of scRNA-seq data for mouse tracheal epithelial cells ($n = 4$ mice, 7,662 cells) (**c**) and HBECs ($n = 3$ donors, 2,970 cells) (**d**) coloured by inferred cell type, with heat maps of lineage-specific genes by biological replicates (rows). Cell numbers are post-quality control. PNEC, pulmonary neuroendocrine cells. Lineage markers for PNECs and brush cells were expressed in rare cells in HBEC cultures, and formed just one human cluster.

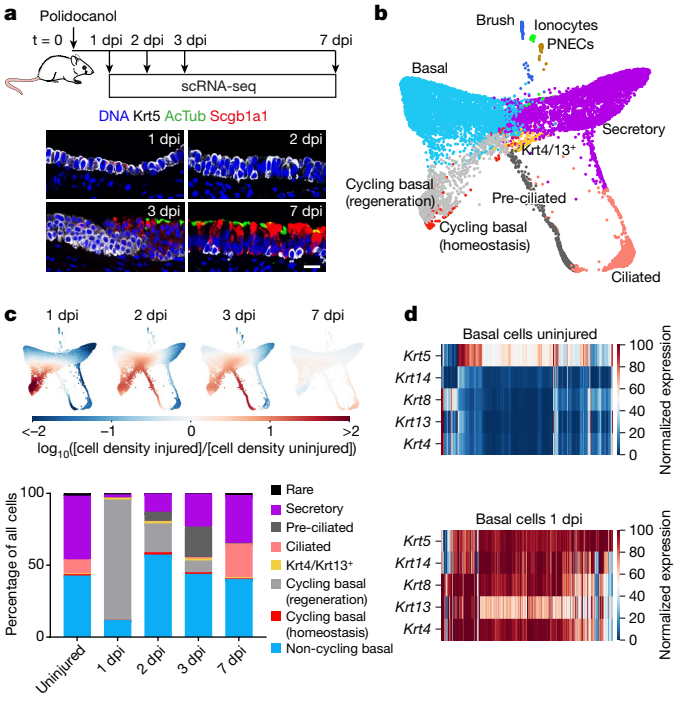


Fig. 2 | Single-cell RNA-seq reveals recovery-specific cell states. a, Mice were administered 2% polidocanol by oral-pharyngeal aspiration, and tracheae were collected 1 ($n = 1$), 2 ($n = 1$), 3 ($n = 1$) and 7 ($n = 3$) dpi for scRNA-seq. Immunofluorescence for basal cells (Krt5) and lumenal markers (AcTub and Scgb1a1) shows that lumenal lineages are shed at 1 dpi ($n = 3$), basal population expands at 2 dpi ($n = 4$), mature lumenal markers are visible at 3 dpi ($n = 3$) and the differentiated epithelium is restored by 7 dpi (compare to Fig. 1b; $n = 3$). **b**, SPRING plot of scRNA-seq data showing cells from uninjured ($n = 7{,}898$) and regenerating ($n = 6{,}265$) mice. Cell states that emerge during regeneration are shown in grey (see Extended Data Fig. 5a, Methods). **c**, Top, enrichment of scRNA-seq cell states compared to uninjured. Bottom, relative abundance of cell types at each time point. Rare class includes ionocytes, brush and PNECs. **d**, Expression patterns of keratin genes in basal cells change between uninjured trachea and 1 dpi. The heat maps show imputed expression counts, with range from the 5th to the 95th percentile. Basal and Krt4/Krt13$^+$ cluster cells shown.

was closely correlated and defined a major axis of heterogeneity in basal and differentiating HBECs (Extended Data Fig. 3b).

Second, in the human single-cell map, we identified a *FOXN4*$^+$ cluster that was highly enriched for the ciliated cell specification factor *FOXJ1* but low for markers of maturation, including the ciliary component *TUBB4B* (Fig. 1d, Supplementary Table 2). Foxn4 is known to drive robust transcription of ciliated genes during multiciliated cell differentiation in *Xenopus*[16], suggesting that this cluster represents a state of multiciliated cell differentiation. We confirmed the existence of this cluster by immunofluorescence, showing that FOXN4 was indeed enriched in a subset of FOXJ1^{high} cells but not in cells containing mature cilia (Extended Data Fig. 4b). Third, in the human data we identified a novel cluster that was enriched for *SLC16A7* (Fig. 1d, Supplementary Table 2), which encodes the monocarboxylate transporter 2 (MCT2) that is involved in acidification of cystic fibrosis HBEC cultures[17], as well as *AIRE*, the gene that drives negative selection of self-reactive T cells in thymic epithelium[18]. This cluster contained the largest number of highly specific genes in the dataset, with a greater percentage of mitochondrial genes. This cluster may reflect cellular stress[19] or may represent a unique antigen-presenting airway epithelial cell.

Finally, a single cluster identified in both mouse and human was enriched for ion transporters and the transcription factors *Foxi1*, *Ascl3* and *Tfcp2l1* (Fig. 1c, d, Supplementary Tables 1, 2). This cluster expressed subunits of the V-ATPase proton pump, which are also expressed in *Foxi1*-expressing ionocytes in the mucociliary epithelium of *Xenopus* larval skin, intercalated cells of the mammalian kidney and in forkhead-related (FORE) cells of the inner ear[20,21]. This cluster was highly enriched for *Cftr*, the gene that encodes a critical chloride channel that is mutated in cystic fibrosis, as well as for genes encoding multiple CLC chloride channels (for example, NKCC1, ClC-Kb), the calcium-activated potassium channel KCNMA1 and members of the Slc9 family of Na^+/H^+ exchangers (Nhe4 in mouse and NHE7 in human). We named these cells pulmonary ionocytes.

To further identify cell states in the conducting airway that may emerge or expand following injury, we performed scRNA-seq and immunofluorescence on regenerating tracheas at 1, 2, 3 and 7 days after polidocanol-induced injury (Fig. 2a). We visualized transcriptomes of these cells using a SPRING graph, expanded to 14,163 cells to reveal detailed changes in epithelial cell states during repair (Fig. 2b). This identified two states specific to injury response (Extended Data Fig. 5a, Methods). The first state appeared at 1 day(s) post-injury (dpi) (Fig. 2b (light grey), c), corresponding to *Krt5*$^+$ basal cells in cycle and co-expressing additional cytokeratins including *Krt14, Krt8* and *Krt4/Krt13*, which were largely non-overlapping in homeostasis (Fig. 2d). The second injury-specific state, which appeared at 2 and 3 dpi, included cells transiting directly from basal to ciliated (Fig. 2b (dark grey), c) rather than differentiating through a secretory progenitor (Fig. 1c). We detected 1,237 genes that varied in expression during multiciliated cell differentiation, including the specification factors *Foxj1*, *Myb* and *Mcidas*[22] (Extended Data Fig. 6, Supplementary Table 4). Early secretory cell states also reappeared at 2 and 3 dpi. By 7 dpi, the relative abundance of cell populations, including rare

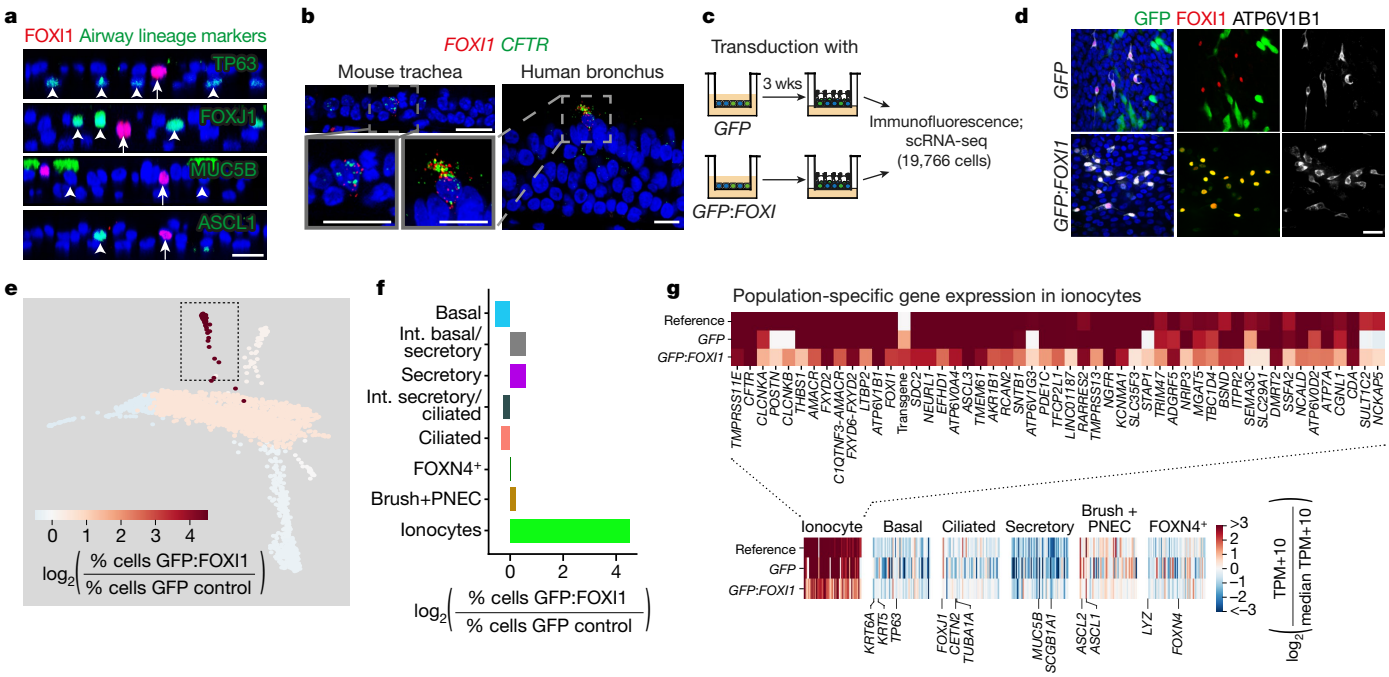


Fig. 3 | FOXI1 specifies a novel cell type, the CFTR-rich 'pulmonary ionocyte'. **a**, Immunofluorescence of FOXI1 (red, arrow), and airway lineage markers (green, arrowheads); TP63 (basal), FOXJ1 (ciliated), MUC5B (secretory) and ASCL1 (PNEC) in differentiated HBEC cultures ($n = 3$ donors). **b**, Fluorescent in situ hybridization in mouse tracheal epithelium ($n = 3$ mice) and human bronchial epithelium ($n = 2$ donors) for *FOXI1* (red) and *CFTR* (green). **c**, HBECs were transduced at seeding with *GFP* or *GFP:FOXI1* lentivirus, differentiated and then profiled by scRNA-seq or analysed by immunofluorescence. **d**, Immunofluorescence for ATP6V1B1 (white) and FOXI1 (red) in HBECs transduced with *GFP* or *GFP:FOXI1* ($n = 4$ experiments from two donors). Scale bars, 20 μm. **e**, **f**, Fold change in fractions of cell states revealed by scRNA-seq in *GFP:FOXI1* versus *GFP*. Heat map values correspond to the ratio of cell numbers from the viral transduction experiments projecting onto each point of the reference HBEC dataset from Fig. 1d. Extended Data Fig. 8e extends this to populations specific to viral transduction. **g**, Ionocytes induced by *GFP:FOXI1* are transcriptionally similar to natural ionocytes, shown by comparing their gene expression in scRNA-seq data from three experimental conditions (reference data from Fig. 1d). The genes shown are markers of each epithelial cell type (bottom), with ionocyte markers shown in detail (top). Genes are normalized to the median expression level across populations observed in a given condition. *ADGRF5* was previously known as *GPR116*.

populations (PNECs, brush cells and pulmonary ionocytes) largely returned to that seen in uninjured tracheae (Fig. 2c, Extended Data Fig. 5b).

Our data open up a range of possible avenues for future research, from the importance of the gene modules defining basal and secretory cell heterogeneity, to the catalogue of potential regulators and components for rare PNECs and brush cells, premature ciliated cells, *Krt4*/*Krt13*$^+$ cells and pulmonary ionocytes. In this study, we focus on the localization, specification and function of the newly identified pulmonary ionocyte.

We first validated the presence of the pulmonary ionocyte population by immunofluorescence. FOXI1 labelled 1–2% of HBECs and, as predicted, was distinct from basal cells (labelled by TP63), secretory cells (MUC5B), ciliated cells (FOXJ1) or neuroendocrine cells (ASCL1) (Fig. 3a). Immunostaining demonstrated apical enrichment of the V-ATPase in FOXI1$^+$ HBECs, similar to what has been shown for other Foxi1$^+$ epithelial lineages[20,21,23], as well as nerve growth factor receptor (NGFR, Extended Data Fig. 4c), confirming the marker gene enrichment identified by scRNA-seq (Extended Data Fig. 1b). Also as predicted, *CFTR* mRNA was highly enriched in *FOXI1*-expressing cells in mouse trachea and primary human bronchial tissue (Fig. 3b), compared to the low expression throughout the epithelium (Extended Data Fig. 7d, e). *FOXI1*$^+$ cells were more concentrated in bronchial gland ducts than in the surface epithelium (Extended Data Fig. 7d), a pattern similar to previously described rare, CFTR^{high} cells[24].

In other proton-secreting cells, *Foxi1* specifies the lineage and regulates expression of V-ATPase subunits[20,21]; therefore, we next asked whether *FOXI1* was sufficient to specify pulmonary ionocytes in HBEC cultures. We performed scRNA-seq of cells transduced with lentivirus expressing *GFP:FOXI1* ($n = 10{,}330$) or *GFP* alone ($n = 9{,}436$) and mapped the data onto the reference HBEC state map (Fig. 3c, e, Extended Data Fig. 8a–e). Cultures transduced with *GFP:FOXI1* had significantly higher numbers of cells classified as ionocytes (23-fold increase, $P < 10^{-50}$ by Fisher's exact test, Fig. 3e, f), with slight reductions in basal and ciliated cells. The resulting ionocytes expressed high levels of exogenous *GFP:FOXI1* (Fig. 3g, Extended Data Fig. 8f), and exhibited the same transcriptional program as unperturbed ionocytes. Moreover, they did not express markers of other cell types (Fig. 3g). Immunostaining (Fig. 3d), quantitative PCR with reverse transcription (RT–qPCR) profiling of marker genes, and RNA in situ hybridization performed on transduced cultures (Extended Data Fig. 7a–c) confirmed these results, indicating that *FOXI1* is sufficient to specify *CFTR*-rich pulmonary ionocytes. *FOXI1* overexpression also led to the appearance of a novel non-ionocyte cell state, possibly resulting from off-target FOXI1 transcriptional activity (Extended Data Fig. 8c–e, Supplementary Table 3).

In *Xenopus* epidermis, ionocytes differentiate from an inner layer of basal cells, and their specification is regulated by Notch signalling[21,23], a pathway that is important in airway basal cell fate[25–27]. The specification of pulmonary ionocytes shows clear similarities. *Foxi1*$^+$ cells first reappeared in the basal cell pool following injury depletion (Extended Data Fig. 5c), and Foxi1 co-localization with the basal cell marker Krt5 transiently increased after injury (46.3% of Foxi1$^+$ cells at 3 dpi compared to 15.4% at steady state; Fig. 4a). This suggests that Foxi1$^+$ cells have a direct basal cell origin, consistent with steady-state lineage tracing studies[11]. Notch target genes are also expressed in pulmonary ionocytes (Extended Data Fig. 9). Treating HBEC cultures with the γ-secretase inhibitor DAPT decreased Notch target gene expression and increased ciliated cell specification, consistent with previous studies in which Notch signalling was modulated in mouse and human airway

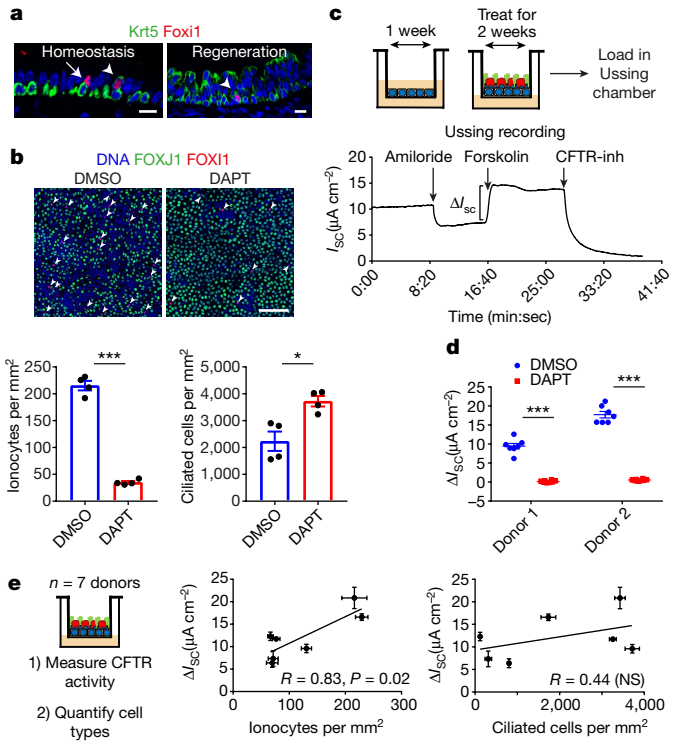


Fig. 4 | Pulmonary ionocytes are a major source of CFTR activity. **a**, Immunofluorescence for Krt5 (green) and Foxi1 (red) in mouse tracheae at homeostasis (left, $n = 3$) or 3 dpi (right, $n = 3$). Arrowhead, Foxi1$^+$Krt5$^+$ cells; arrow, Foxi1$^+$Krt5$^-$ cell. Scale bars, 20 μm. **b**, Immunofluorescence and quantification for ionocytes (FOXI1$^+$, red, arrowheads) and ciliated cells (FOXJ1$^+$, green) in HBEC cultures treated with DMSO or DAPT. Scale bar, 100 μm; $n = 4$ experiments from one donor. $^*P = 0.01$, $^{***}P = 1.1 \times 10^{-6}$ by two-tailed t-test. **c**, HBECs were treated with DMSO or DAPT upon ALI culture. After differentiation (2–3 weeks), cultures were loaded into Ussing chambers and short-circuit current (I_{sc}) was recorded during addition of amiloride, forskolin, and a CFTR inhibitor, CFTR(inh)-172. A representative tracing from donor 1 ($n = 11$) is shown. **d**, Change in short-circuit current (ΔI_{sc}) in response to forskolin measured in DMSO ($n = 7$ cultures per donor) or DAPT-treated cultures ($n = 8$ cultures per donor). $^{***}P < 1 \times 10^{-8}$ by two-tailed t-test. **e**, Donor mean ΔI_{sc} in response to forskolin plotted against mean number of ionocytes (FOXI1$^+$) or ciliated cells (FOXJ1$^+$) ($n = 7$). All data are mean ± s.e.m. R, Pearson correlation with associated P value. NS, not significant.

cultures[28] (Extended Data Fig. 10a). DAPT treatment also significantly decreased the number of ionocytes (Fig. 4b). Treating HBEC cultures with antibodies against individual Notch receptors also reduced ionocyte numbers (Extended Data Fig. 10b).

We next investigated the functional importance of the high level of *CFTR* expression in pulmonary ionocytes. Ciliated cells have been hypothesized to be the major source of CFTR in the proximal airway[29], but we found little to no *CFTR* expression in *FOXJ1*$^+$ ciliated cells (Supplementary Table 2, Extended Data Fig. 7e). To examine CFTR activity in the proximal airway epithelium, we recorded CFTR-mediated ion transport in HBEC cultures using Ussing chambers[30] (Fig. 4c). DAPT-treated cultures, which reduce the number of ionocytes and increase the number of ciliated cells, had significantly lower CFTR activity in response to forskolin (measured as short-circuit current, I_{sc}; Fig. 4d). We also used natural variation between donors to assess the sensitivity of CFTR activity to changes in ionocyte numbers versus changes in ciliated cell numbers. Ussing experiments and cell-type quantification in cultures derived from seven different donors showed that CFTR activity was positively correlated with ionocyte number (Pearson's $R = 0.83$, $P = 0.02$) and not correlated with ciliated cell number ($R = 0.44$, $P = 0.32$), with ionocytes explaining 60% of the mean channel current compared to just 4% for ciliated cells (after multivariate regression) (Fig. 4e). These data suggest that ionocytes are a major source of CFTR activity in airway epithelium despite representing only 1–2% of epithelial cells.

In this study, we applied large-scale single-cell profiling to conduct an unbiased investigation of the composition of the proximal airway epithelium during homeostasis and regeneration. In doing so, we unexpectedly identified the pulmonary ionocyte, a rare cell type that appears to be a major, possibly dominant, source of CFTR activity in airway epithelium. This cell type shows co-enrichment of the proton-secreting V-ATPase and the anion-secreting CFTR channel, suggesting a role in luminal pH regulation that could be relevant for the pathology of cystic fibrosis[31]. The role of these CFTR-rich cells in airway physiology and disease remains to be elucidated, but it is likely that their identification will better inform future therapeutics for cystic fibrosis. Finally, our study provides a comprehensive atlas of genes and pathways with potential roles in promoting differentiation and repair, delineates the cell types, transcriptional profiles and trajectories present in the proximal airway in both homeostatic and regenerating tissues, and offers a baseline for future profiling of disease states.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0394-6.

Received: 31 July 2017; Accepted: 21 June 2018;
Published online 1 August 2018.

1. Rock, J. R., Randell, S. H. & Hogan, B. L. M. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis. Model. Mech.* **3**, 545–556 (2010).
2. Hogan, B. L. M. et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15**, 123–138 (2014).
3. Hong, K. U., Reynolds, S. D., Watkins, S., Fuchs, E. & Stripp, B. R. In vivo differentiation potential of tracheal basal cells: evidence for multipotent and unipotent subpopulations. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **286**, L643–L649 (2004).
4. Rock, J. R. et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106**, 12771–12775 (2009).
5. Pardo-Saganta, A. et al. Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell* **16**, 184–197 (2015).
6. Van Lommel, A. Pulmonary neuroendocrine cells (PNEC) and neuroepithelial bodies (NEB): chemoreceptors and regulators of lung development. *Paediatr. Respir. Rev.* **2**, 171–176 (2001).
7. Reid, L. et al. The mysterious pulmonary brush cell: a cell in search of a function. *Am. J. Respir. Crit. Care Med.* **172**, 136–139 (2005).
8. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
9. Gray, T., Guzman, K., Davos, W., Abdullah, L. & Nettesheim, P. Mucociliary differentiation of serially passaged normal human tracheobronchial epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **14**, 104–112 (1996).
10. Tata, P. R. & Rajagopal, J. Plasticity in the lung: making and breaking cell identity. *Development* **144**, 755–766 (2017).
11. Montoro, D. T. et al. A revised airway hierarchy includes CFTR-expressing ionocytes. *Nature* https://doi.org/10.1038/s41586-018-0393-7 (2018).
12. Weinreb, C., Wolock, S. & Klein, A. M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2017).
13. Cole, B. B. et al. Tracheal basal cells: a facultative progenitor cell pool. *Am. J. Pathol.* **177**, 362–376 (2010).
14. Watson, J. K. et al. Clonal dynamics reveal two distinct populations of basal cells in slow-turnover airway epithelium. *Cell Reports* **12**, 90–101 (2015).
15. Reynolds, S. D., Reynolds, P. R., Pryhuber, G. S., Finder, J. D. & Stripp, B. R. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. *Am. J. Respir. Crit. Care Med.* **166**, 1498–1509 (2002).
16. Campell, E. P., Quigley, I. K. & Kintner, C. Foxn4 promotes gene expression required for multiple motile cilia formation. *Development* **143**, 4654–4664 (2016).
17. Garnett, J. P. et al. Hyperglycaemia and *Pseudomonas aeruginosa* acidify cystic fibrosis airway surface liquid by elevating epithelial monocarboxylate transporter 2 dependent lactate-H$^+$ secretion. *Sci. Rep.* **6**, 37955 (2016).
18. Anderson, M. S. & Su, M. A. AIRE expands: new roles in immune tolerance and beyond. *Nat. Rev. Immunol.* **16**, 247–258 (2016).
19. Galluzzi, L., Kepp, O. & Kroemer, G. Mitochondria: master regulators of danger signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 780–788 (2012).
20. Vidarsson, H. et al. The forkhead transcription factor Foxi1 is a master regulator of vacuolar H$^+$-ATPase proton pump subunits in the inner ear, kidney and epididymis. *PLoS ONE* **4**, e4471 (2009).

21. Quigley, I. K., Stubbs, J. L. & Kintner, C. Specification of ion transport cells in the *Xenopus* larval skin. *Development* **138**, 705–714 (2011).
22. Spassky, N. & Meunier, A. The development and functions of multiciliated epithelia. *Nat. Rev. Mol. Cell Biol.* **18**, 423–436 (2017).
23. Dubaissi, E. & Papalopulu, N. Embryonic frog epidermis: a model for the study of cell–cell interactions in the development of mucociliary disease. *Dis. Model. Mech.* **4**, 179–192 (2011).
24. Engelhardt, J. F. et al. Submucosal glands are the predominant site of CFTR expression in the human bronchus. *Nat. Genet.* **2**, 240–248 (1992).
25. Rock, J. R. et al. Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell* **8**, 639–648 (2011).
26. Tsao, P. et al. Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development* **136**, 2297–2307 (2009).
27. Danahay, H. et al. Notch2 is required for inflammatory cytokine-driven goblet cell metaplasia in the lung. *Cell Reports* **10**, 239–252 (2015).
28. Guseh, J. S. et al. Notch signaling promotes airway mucous metaplasia and inhibits alveolar development. *Development* **136**, 1751–1759 (2009).
29. Kreda, S. M. et al. Characterization of wild-type and ΔF508 cystic fibrosis transmembrane regulator in human respiratory epithelia. *Mol. Biol. Cell* **16**, 2154–2167 (2005).
30. Li, H., Sheppard, D. N. & Hug, M. J. Transepithelial electrical measurements with the Ussing chamber. *J. Cyst. Fibros.* **3**, 123–126 (2004).
31. Shah, V. S. et al. Airway acidification initiates host defense abnormalities in cystic fibrosis mice. **351**, 503–507 (2016).

Acknowledgements We thank members of the Jaffe and Klein labs, and of the Tissue Repair hub for helpful discussions; C. Weinreb for help with implementing SPRING and data analysis; C. Bauer and K. Feldman for help obtaining human bronchial tissue; P. Capodieci, K. Wetzel and J. Judge for help with tissue processing; K. Mansfield, B. Suchomel and C. Saravanan for help with histology; K. Coote and R. Valdez Misiolek for help with electrophysiology; N. Kirkpatrick, A. Szilvasi and D. Ahern-Ridlon for help with flow cytometry and imaging; J. Reece-Hoyes and S. An for help with lentiviral design; the Single Cell Core Facility at Harvard Medical School for inDrop reagents; the Bauer Core Facility for sequencing; T. Bouwmeester and J. Porter for support and input through the course of this work; R. DeBerardinis for reviewing the manuscript; and B. Ward for editorial assistance. L.W.P. would like to thank R. N. Plasschaert for thoughtful discussion. A.M.K. received support from a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, an Edward Mallinckrodt Jr. Foundation Grant. R.Z. received support from the Lithuanian Education Exchanges Support Foundation. A.M.K. and R.Z. are supported by NCI grant R33CA212697-01.

Reviewer information *Nature* thanks I. Amit and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.M.K. and A.B.J. conceived the project and performed the initial pilot experiments. L.W.P., R.Z., G.R., A.M.K. and A.B.J. designed experiments and L.W.P. and R.Z. performed most experiments. J.K. and G.R. sequenced cDNA libraries. R.Z, V.S. and A.M.K. performed scRNA-seq analysis. L.W.P. performed histology, lentiviral studies, Notch pathway modulation and electrophysiology. R.C.-W. performed animal work including polidocanol dosing and trachea isolation. L.W.P., R.Z., A.M.K. and A.B.J. wrote the manuscript. All authors read and reviewed the manuscript.

Competing interests L.W.P., R.C.-W., J.K., G.R., and A.B.J. are employees of Novartis Institutes for BioMedical Research. V.S. is an employee of Sanofi Aventis. A.M.K. is a founder and SAB member of 1Cell-Bio.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0394-6.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0394-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.M.K. or A.B.J.

METHODS

HBEC culture and Notch inhibition. Primary human bronchial epithelial cells (HBECs) from normal donors aged 3–42 were obtained from Lonza (CC-2540; Lot 221175, 323353, 429581, 105104, 621713, 312626 and 441099) and were expanded twice with growth medium (500 ml BEGM medium (Lonza, CC-3171), 1 SingleQuots kit (Lonza, CC-4175)) in T75 flasks. After expansion, HBECs were seeded on 12-well Transwell plates (Corning, 3460) at a density of 83,000 cells per Transwell. The cells were cultured in differentiation medium (250 ml BEGM medium, 250 ml DMEM medium (Thermo Fisher, 11965092), 1 SingleQuots kit) on both apical and basal sides of Transwells for the first 7 days. Then, medium was removed from apical side, and cells were cultured for another two weeks at an ALI condition. Cells were used for analysis after culture at ALI for 14 days and no more than 28 days.

Notch signalling was inhibited by adding 3.3 μM DAPT or DMSO to differentiation medium when HBECs were cultured at ALI. Notch antibodies against receptors NOTCH1, NOTCH2 and NOTCH3 and control IgG antibody were described previously[27,32–34] and were added at a concentration of 10 μg/ml to HBECs upon culture at ALI.

Single-cell dissociation. HBECs were collected after 15 days of ALI culture using 0.05% trypsin–EDTA (Thermo Fisher, 25300054). Cells were then pelleted at 300*g* for 5 min, resuspended in PBS and filtered through a 20-μm strainer (PluriSelect, 43-50020-03). Cells were counted on a haemocytometer and Optiprep (Sigma-Aldrich, D1556) was added to achieve a final concentration of 15% and 75,000 cells/ml.

C57/BL6 male and female mice from the Jackson Laboratory aged 6–8 weeks were used for all studies. Animals were handled in accordance with Novartis Institutes for Biomedical Research Animal Care and Use Committee protocols and regulations. Mice were housed in a temperature- and humidity-controlled animal facility with ad libitum access to food and water and acclimated for at least 3 days before experimental manipulation. For single-cell isolation for scRNA-seq, tracheas were dissected and opened longitudinally in Ham's F12 (Life Technologies, 11765-054) plus 1% penicillin–streptomycin on ice. Each trachea was individually placed in a 15-ml conical tube with 5 ml 1.5 mg/ml Pronase (Roche, 10165921001) in Ham's F12 plus 1% penicillin–streptomycin and incubated for 18 h at 4 °C. Five hundred microlitres FBS was added to inactivate pronase and conical tubes were vigorously inverted to dislodge cells. Each trachea was transferred twice to a 15-ml conical tube containing Ham's F12 plus 1% penicillin–streptomycin plus 10% FBS and then inverted. Medium from each of the three tubes was pooled and cells were pelleted at 400*g* for 10 min at 4 °C. Cells were resuspended in 500 μl DNase solution (0.5 mg/ml DNase (Sigma-Aldrich, DN25), 10 mg/ml BSA in Ham's F12 + 1% penicillin–streptomycin), incubated on ice for 5 min and then pelleted at 400*g* for 10 min at 4 °C. Cells were then washed twice in Ham's F12 1% penicillin–streptomycin 10% FBS and then resuspended in PBS + 0.02% BSA. Cells were diluted to 90,000 cells/ml in 15% Optiprep + 0.02% BSA in PBS for scRNA-seq.

Single-cell transcriptome barcoding in drops and library preparation for Illumina sequencing. For scRNAseq, we used inDrops[8] following the protocol previously described[35] with the modifications summarized in Supplementary Table 5. In brief, dissociated single cells were co-encapsulated into 3–4-nl droplets together with hydrogel beads carrying barcoding reverse transcription primers. Following a reverse transcription in droplets, the emulsion was broken and the bulk material was taken through the following steps: (i) second strand synthesis; (ii) linear amplification by in vitro transcription (IVT); (iii) amplified RNA fragmentation; (iv) reverse transcription; (v) PCR. A subset of the GFP:FOXI1 libraries were processed using small variations on the published protocol, including a different reverse transcriptase and exclusion of HinfI digestion. The resulting libraries were sequenced either on a HiSeq or Nextseq Illumina platform in paired-end mode to a length of 2 × 100 or 2 × 76 base pairs (see Supplementary Table 5). Images from the instrument were processed using the manufacturer's software to generate FASTQ sequence files. Read quality was assessed by running FASTQC (v.0.10.1).

Obtaining transgene counts in cells transduced with lentivirus. The single cell RNA-seq method we used allows the detection of transcript sequences up to ~1 kb upstream of the polyA tail. Both *GFP* and *GFP:FOXI1* transcripts share the same 1.3-kb-long sequence upstream of the transcription termination and polyadenylation site within the lentiviral 3′ long terminal repeat (LTR). This 1.3-kb sequence, which is part of the plenti6/V5-DEST.NGFP Gateway scaffold, was added to the reference transcriptome to identify transgene counts. Notably, 'transgene' refers to either *GFP:FOXI1* or *GFP*, depending on the dataset. In Fig. 3g, the transgene was added manually to the heat map (top), as was *FOXJ1* (bottom heat map), a canonical marker of multiciliated cells that fails to appear as a unique marker gene because of its expression in the *FOXN4*$^{+}$ cluster.

Single-cell RNA-seq data analysis. *Processing of sequencing reads.* To generate per-cell gene expression counts from raw sequencing reads, we used an updated and publicly available version (https://github.com/indrops) of the custom sequencing data-processing pipeline described[8]. Parameters used with the indrop.py pipeline are specified in .yaml files provided as Supplementary Files 1 and 2. In brief, raw reads (FASTQ) were filtered for sequencing quality and expected structure, sorted based on barcodes sequences (reads derived from the transcriptome of the same cell carry the same barcode), and aligned to either mm10 or hg19 cDNA reference with separately added mitochondrial cDNA sequences. To quantify gene expression while correcting for amplification biases, we made use of unique molecular identifiers (UMIs) introduced during reverse transcription in drops. The output of low-level processing is a genes × cells expression matrix.

Single-cell data cleanup and normalization. To ensure high-quality data for further analysis, we filtered out cells with few counts and a high mitochondrial gene fraction. Thresholds were selected by visually inspecting histograms of counts per cell and mitochondrial fraction per cell for each biological sample separately. With human samples, a mitochondrial fraction threshold of 25% and total count thresholds of 1,500, 1,500 and 2,000 were used for donors 1, 2 and 3, respectively. The same thresholds as for donor 1 were used in the GFP/GFP:FOXI1 overexpression experiment. For mouse, a mitochondrial fraction threshold of 20% and total count thresholds of 1,500 were used for all datasets. Initial visualization and clustering (see below) revealed that a small fraction of mouse cells (<4%) formed well-separated clusters characterized by a strong immune gene signature. These cells were excluded from further analysis. For uninjured mouse data, we applied an additional clean up step to remove cell doublets, which can occur rarely owing to incomplete cell dissociation or owing to two cells occasionally entering the same microfluidic barcoding droplet. In brief, a decoy training set of simulated doublets is generated by randomly combining single-cell transcriptomes from the dataset. This decoy training set is used to train a *k*-nearest neighbour classifier. Cell transcriptomes classified as in silico doublets were excluded from further analysis. The detailed method will be published elsewhere. Data was then normalized by the total counts per cell, as described[8], with the following modification: to calculate the normalization factor (total counts per cell) we excluded any gene with expression level >5% of total counts in at least one cell.

Data visualization using SPRING and clustering. To visualize the high-dimensional gene expression data, we applied SPRING[12], a method for building a *k*-nearest neighbours (*k*NN) graph of cells and representing it in 2D using a force-directed layout. Clusters were identified by applying spectral clustering on the same adjacency matrix as used for SPRING (implementation in python, sklearn.cluster. SpectralClustering(affinity = 'precomputed', assign_labels = 'discretize')). Clusters were assigned labels (for example, secretory, basal) based on marker gene expression. In the SPRING plot of human data (Fig. 1), clusters representing intermediate states with no unique gene expression are shown in grey.

Cell population-specific gene identification (Fig. 1). To be considered as specific to population *i*, a gene had to satisfy the following criteria: (a) be expressed at significantly higher level in population *i* compared to all other cells as determined by a two-sided permutation test using the difference in sample means as the test statistic (false discovery rate (FDR) <5%). To be considered for statistical testing, a gene had to be detected in a least 1% of cells on either side of the comparison. (b) Average expression >50 transcripts per million (TPM) in population *i*. (c) Average expression in population *i* at least 1.5-fold higher than in any other population (that is, maximum-to-second-maximum ratio >1.5). A pseudo value of 10 TPM was added before division. (d) Be maximum in population *i* for 4/4 (mouse) or 2/3 (human) of the biological replicates.

Figure 1c, d shows the expression level of the top 50 such hits ordered by decreasing maximum-to-second-maximum ratio. For each gene, 100% was set at the maximum expression per cluster (average of all replicates). The colour was saturated at 20% (low) and 100% (high). Detailed gene lists are provided as Supplementary Tables 1–3.

For Extended Data Fig. 1, transcription factor lists were obtained from animaltfDB[36], and GO terms GO:0016301 and GO:0009986, including any descendent terms, were used for kinases and surface molecules, respectively.

Identification of correlated gene modules within basal and secretory cells. To characterize the heterogeneity within basal and secretory cells, we identified modules of correlated genes. For mouse, we performed the following steps (see also Extended Data Figs. 2, 3): (a) select basal cells (same procedure for secretory); (b) identify 3,000–5,000 most variable genes; (c) calculate gene–gene rank correlation; (d) retain genes that have $r > 0.2$ with at least 4 other genes—in mouse, *Krt14* did not meet this criterion, and was therefore was included manually.

Heat map rows and columns (Extended Data Figs. 2, 3) were hierarchically clustered (distance defined as $1-r_{\text{Spearman}}$, Ward linkage). For human data, we first considered the basal, secretory and intermediate-state cells collectively to identify two main modules of anti-correlated genes (Extended Data Fig. 3). From there, we selected genes specific to basal and recalculated gene–gene correlation but within basal cells only. The same was performed with secretory cells.

Smoothing (data imputation). Smoothing was carried out in Fig. 2c, d, and Extended Data Figs. 5a, 8c, 9. All data shown in other figures is not subject to

smoothing/imputation. Data smoothing, or equivalently imputation, was carried out using a graph diffusion approach on the *k*-nearest neighbour graph *G* defined above by SPRING. *G* is an unweighted undirected graph. The smoothing operation replaces a scalar quantity x_i on node *i* of the graph, for example, raw expression level of a gene, with a smoothed value $x^{(s)} = O_S x$, in which the smoothing operator is $O_S = e^{L\beta}$ and *L* is the random walk graph Laplacian of *G*. The smoothing operator accepts a single parameter, β, which determines the kernel size, that is, the extent of smoothing. This parameter is equivalent in physical terms to diffusion time: longer times lead to broader diffusion. For all plots shown, we used $\beta = 1$.

Analysis of cell density changes relative to uninjured. To visualize which cell populations are enriched at a given time point relative to uninjured (Fig. 2) the following was performed for every time point *t* of mouse recovery data: (a) get every cell from *t* to vote for its 10 nearest neighbours among all mouse cells and count votes. (b) Smooth vote counts on the graph (see previous section for smoothing). Smoothed vote counts are a proxy for the density of cells from time point *t* on the graph (see also the two left-most plots of Extended Data Fig. 5a). (c) Normalize the total vote count to 1. (d) Divide the density at time point *t* by the density of cells in uninjured.

Identification of recovery-specific cell populations. The procedure is summarized in Extended Data Fig. 5a. To identify recovery-specific cell populations in the SPRING plot combining all mouse data (populations in grey in Fig. 2b), we first performed steps (a) and (b) described in the previous paragraph to determine the density of injured cells on the graph. Next, a threshold of 25 smoothed counts was selected by visual inspection of the distribution of votes, and cells receiving fewer than 25 votes were considered depleted in uninjured (that is, recovery-specific). Recovery-specific cells were split into two clusters by spectral clustering, and labels were assigned based on characteristic gene expression. Cells from the mouse recovery time course experiment that were not recovery-specific inherited the label of their single nearest neighbour in uninjured mouse data (Euclidean distance in principal component space of most variable genes).

Analysis of recovery-specific trajectory from basal to ciliated. The procedure is also summarized in Extended Data Fig. 6. Six hundred and nine recovery-specific cells from 24–72 h post-injury, and forming a continuum between basal and ciliated cells, were manually selected on the SPRING plot and used for population balance analysis (PBA), a method developed in our laboratory for describing differentiation trajectories[37]. For this analysis, 'source' and 'sink' cell populations were defined as the basal and multiciliated tips of the cell *k*NN graph, respectively. Cells were then ordered on the graph by the diffusion 'potential' parameter of PBA (a measure pseudotime of progression from source to sink). To smooth the gene expression of individual cells, a moving average with window size of 100 cells was calculated.

Identification of differentially expressed genes along the basal-to-ciliated trajectory. Temporally varying genes were identified using a previous method[38] with minor changes. Before statistical testing, the following filters were applied on the full gene list considering only the 609 cells forming the basal-to-ciliated trajectory: (a) expression level: at least 3 normalized counts in at least 3 cells; (b) variable: Fano factor >1.

Notably, none of these filters considers the cell ordering. For each gene *i* of the surviving 4,561 genes, a statistic *t* was calculated: $t_{i,\text{observed}} = m_{i,\max} - m_{i,\min}$, in which m_i is a vector with the expression level of gene *i* in the 510 average cells after application of a moving average over cells ordered using PBA. The procedure was repeated on shuffled cells for multiple permutations, each time resulting in a $t_{i,\text{random}}$ value. The one-sided *P* value for gene *i* was defined as the fraction of times $t_{i,\text{observed}} \geq t_{i,\text{random}}$. To account for multiple hypothesis testing, the false discovery rate was controlled at 5% using the Benjamini–Hochberg procedure. For each of the 4,561 genes used in the permutation test, we also calculated the maximum fold change defined as: $\text{FC}_{\max} = \frac{m_{i,\max} + 100\text{TPM}}{m_{i,\min} + 100\text{TPM}}$. One thousand two hundred and thirty-seven genes with $\text{FC}_{\max} \geq 2$ and FDR ≤ 5% were considered differentially expressed along the basal-to-ciliated trajectory.

Polidocanol-induced injury. Polidocanol-induced injury was performed as previously described[39]. In brief, mice were anaesthetized and delivered one dose of 13.5 μl 2% polidocanol or PBS vehicle control by oral-pharyngeal aspiration to induce injury. Tracheae were collected at 1, 2, 3 and 7 days following injury for scRNA-seq or for fixation and immunofluorescence.

Immunofluorescence, microscopy and cell counting. For RNAscope and immunofluorescence of paraffin-embedded sections, mouse tracheae were dissected under sterile conditions and HBEC Transwell cultures were isolated using 8-mm biopsy punch (Integra Miltex, 33-37). Primary human bronchial tissue was obtained through the International Institute for the Advancement of Medicine. All tissues were immediately fixed in 10% normal buffered formalin for 18–24 h at room temperature then transferred to PBS and kept at 4 °C until paraffin embedding.

For immunofluorescence of mouse tracheae, 5-μm sections were baked and deparaffinized using standard procedures. After antigen retrieval using pH6 citrate buffer (Abcam), sections were rinsed in PBS and blocked in 10% normal goat serum (NGS) or 10% normal donkey serum (NDS) for 30 min at room temperature. Primary antibody was added overnight at 4 °C. Sections were washed 3× in PBS for 5 min each, and secondary antibody was added for 1 h at room temperature and sections were again rinsed in PBS, followed by Hoechst (1:1,000) for 30 s. For RNAscope, 5-μm sections were prepared according to RNAscope procedures for multiplex fluorescent assay (Advanced Cell Diagnostics, 320850) or dual chromogenic assay (322430). RNAscope probes used were *FOXI1* (476351, 483021), *CFTR* (603291, 483011), and *FOXJ1* (430921). Mounting medium and coverslip were applied and slides were stored at 4 °C for immunofluorescence or room temperature for chromogenic in situ hybridization. For immunofluorescence of whole-mount HBEC Transwell cultures, cells were fixed in 4% paraformaldehyde for 30 min RT, washed 3× 10 min in IF buffer (130 mM NaCl, 7 mM Na_2HPO_4, 3.5 mM NaH_2PO_4, 7.7 mM NaN_3, 0.1% bovine serum albumin, 0.2% Triton X-100, and 0.05% Tween- 20), blocked in 10% NGS IF buffer, stained in primary antibody diluted in 10% NGS IF buffer overnight at 4 °C, washed 3× 20 min in IF buffer, counterstained in secondary antibody diluted in 10% NGS IF buffer plus 1:5,000 Hoechst for 1 h at room temperature, washed 3× 20 min in immunofluorescence buffer, and washed 2× in PBS before mounting. The following antibodies were used: rabbit anti-FOXI1 (1:200, Sigma-Aldrich HPA071469), mouse anti-FOXI1 (1:100, Origene TA800146), goat anti-FOXI1 (1:200 Abcam ab20454) rabbit anti-ATP6V1B1 (1:100, Sigma-Aldrich HPA031847), mouse anti-acetylated α-tubulin (1:1,000 Sigma-Aldrich T6793), rabbit anti-Scgb1a1 (1:200, Millipore 07-623), rabbit anti-MUC5B (Santa Cruz sc-20119), rabbit anti-FOXJ1 (1:200, Sigma-Aldrich HPA005714), mouse anti-FOXJ1 (EBioscience, 1:200, 14-9965-80), rabbit anti-FOXN4 (Sigma-Aldrich HPA050018), mouse anti-ASCL1 (1:00, Beckton-Dickinson 556604), mouse anti-Krt4 (1:100, abcam ab9004), rabbit anti-Krt4 (1:100 Proteintech 16572-1-AP), rabbit anti-Krt5 (1:250, abcam ab52635), chicken anti-Krt5 (1:1,000, BioLegend 9059), mouse anti-NGFR (1:200, ThermoFisher, MA1-18418) and chicken anti-Krt8 (1:200, abcam ab107115). Secondary antibodies used were Alexa Fluor 488, 568, 647 (Life Technologies) at 1:500.

Fluorescent images were collected on a confocal microscope (Axiovert 200; Carl Zeiss), with a 40× objective (Zeiss, Plan-Apochromat 40×/1.3 Ph3 M27), a Yokogawa CSU-X1 spinning disc head, and an electron-multiplying charge-coupled device camera (Evolve 512; Photometrics). Scale bars were added, and images were processed using Zen Blue software (Zeiss) and Photoshop (Adobe). FOXI1^+ and FOXJ1^+ cells were counted using ImageJ software. Chromogenic signals were acquired using a Nuance FX multispectral imaging system (Perkin Elmer) with an Olympus BX61 microscope interfaced with a liquid crystal based camera and tunable filter from 420 nm to 720 nm at 20-nm intervals. Spectral components were unmixed and pseudo-coloured for individual channels.

Lentivirus production. For overexpression, *FOXI1* (GeneID 2299) was cloned into the plenti6/V5-DEST.NGFP Gateway vector, which was generated by transferring the N-EmGFP ORF from pcDNA6.2/N-EmGFP-DEST (Thermo Fisher, Cat# V35620) into pLenti6/V5-DEST (Thermo Fisher, V49610). Lentiviral packaging 4×10^6 293T cells were seeded in a 100-mm poly-D-lysine-coated dish (Corning BioCoat, 356469) one day before transfection with 14 ml of cell growth medium (DMEM (Thermo Fisher, Cat# 11965092), 10% FBS (Clontech 631106), 2mM L-glutamine (Invitrogen 25030), 0.1mM MEM Non Essential Amino Acids (Invitrogen 11140), and 1mM sodium pyruvate MEM (Invitrogen 11360)). For transfection, 7 μg of packaging plasmid DNA (ViraPower lentiviral Packaging Mix, Thermo Fisher K497500) was mixed with 5 μg of expression construct DNA and 36 μl Fugene6 (Promega, E2691). OptiMEM (Thermo Fisher, 31985062) was then added the mixture to a total volume of 800 μl. 293T cells were incubated with the transfection reagent mixture for 24 h before the growth medium was refreshed. At 72 h after transfection, virus was collected, and frozen for future experiments. Packaged virus was added to HBEC cultures 1 h after cell seeding and then removed at feeding the following day.

Flow cytometry and cell sorting. Cells were collected using 0.05% Trypsin-EDTA (Thermo Fisher, 25300054), pelleted at 300*g* for 5 min, suspended in 2% FBS DMEM with EDTA and filtered through a 40-μm strainer before being analysed by flow cytometry or cell sorting. RNA was extracted with Trizol (Invitrogen, 15596026) or RNeasy Mini Kit (Qiagen, 74106). cDNA was synthesized from 1 μg of RNA with qScript XLT cDNA Super Mix kit (Quanta Biosciences, 95161-100). Quantitative PCR (qPCR) was carried out using FastStart Universal Probe Master kit (Roche, 04914058001) with 40 ng of cDNA per reaction. Taqman probe sequences used for qPCR (Applied Biosystems) were: *FOXI1*, Hs00201827_m1; *FOXJ1*, Hs00230964_m1; *P63*, Hs00978340_m1; *GAPDH*, Hs99999905_m1; *CFTR*, Hs00357011_m1; *ATP6V1B1*, Hs00266092_m1; *ITGA6*, Hs01041011_m1; *DNAI2*, Hs01001544_m1; *SCGB1A1*, Hs00171092; *MUC5B*, Hs00861588_m1; *NRARP*, Hs01104102_s1; *HES5*, Hs01387464_g1; *HES1*, Hs00172878_m1; *MUC5AC*, Hs01365601_m1.

Short-circuit current (I_{sc}) measurements in Ussing Chambers. For Ussing studies, HBECs were cultured in 6-well Snapwell plates (Corning, 3801) at a seeding density

of 83,000 cells per well. Snapwell inserts containing differentiated HBECs were then mounted in chambers bathed in Buffer (Kreb's Ringer solution; 400 ml H_2O, 25 ml 2.4M NaCl, 25 ml 0.5M $NaHCO_3$, 25 ml 66.6 mM KH_2PO_4 + 16.6 mM K_2HPO_4, 25 ml 24 mM $CaCl_2$ + 24 mM $MgCl_2$, 0.9 g dextrose). Amiloride (Sigma, A9561) was added apically at 10 μM to inhibit Na^+ absorption, then forskolin (Sigma, F6886) was added apically at 20 μM to stimulate cAMP and finally, CFTR-172 (Sigma-Aldrich, C2992) inhibitor was added apically and basally at 30 μM. Under these conditions, cAMP-stimulated I_{sc} due to addition of forskolin could be attributed to CFTR-mediated Cl^- secretion from basolateral to apical solution.

Statistical analysis. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

The standard error of the mean was calculated from the mean of at least three independent HBEC cultures. The Student's *t*-test (unpaired, two-tailed) was used to compare data between groups, with a *P* value of less than 0.05 considered significant.

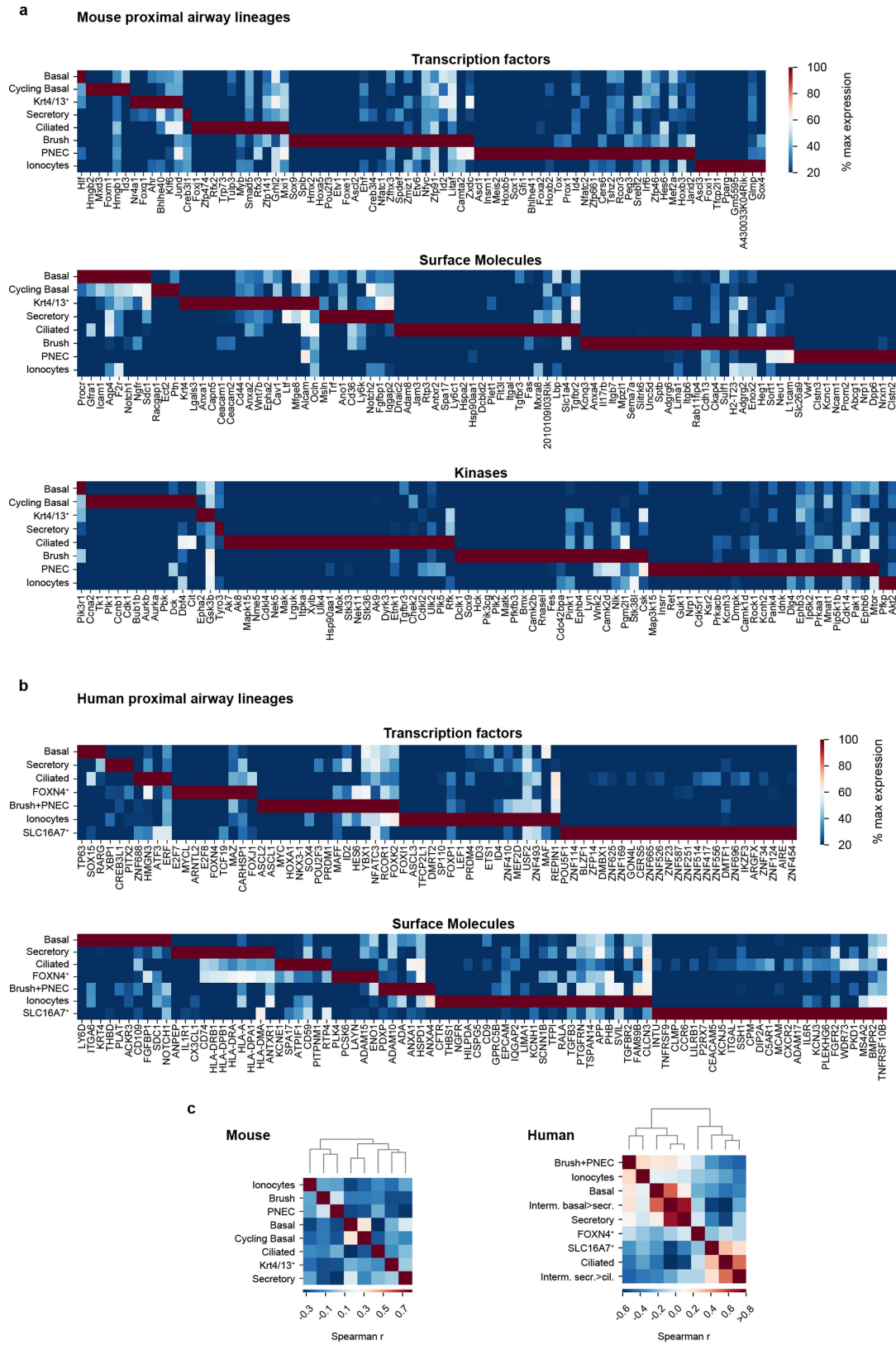
Pearson correlation and its associated *P* value between ΔI_{sc} and $FOXI1^+$ or $FOXJ1^+$ cell number per mm^2 was calculated using the MATLAB corr function. Multivariate regression was carried out using the MATLAB fitlm function. Sensitivity was defined as the fractional change in I_{sc} induced by a fractional change in FOXI1 (x_1) or FOXJ1 (x_2) cell number per mm^2, at the ΔI_{sc} value across all samples, estimated from the slope and intercept of multivariate regression as $S_i = (dI_{sc}/dn)/(I_{sc}/n) = [\text{slope}_i \times x_i]/I_{sc}$ with $i = 1$ and 2, for FOXI1 and FOXJ1, respectively.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

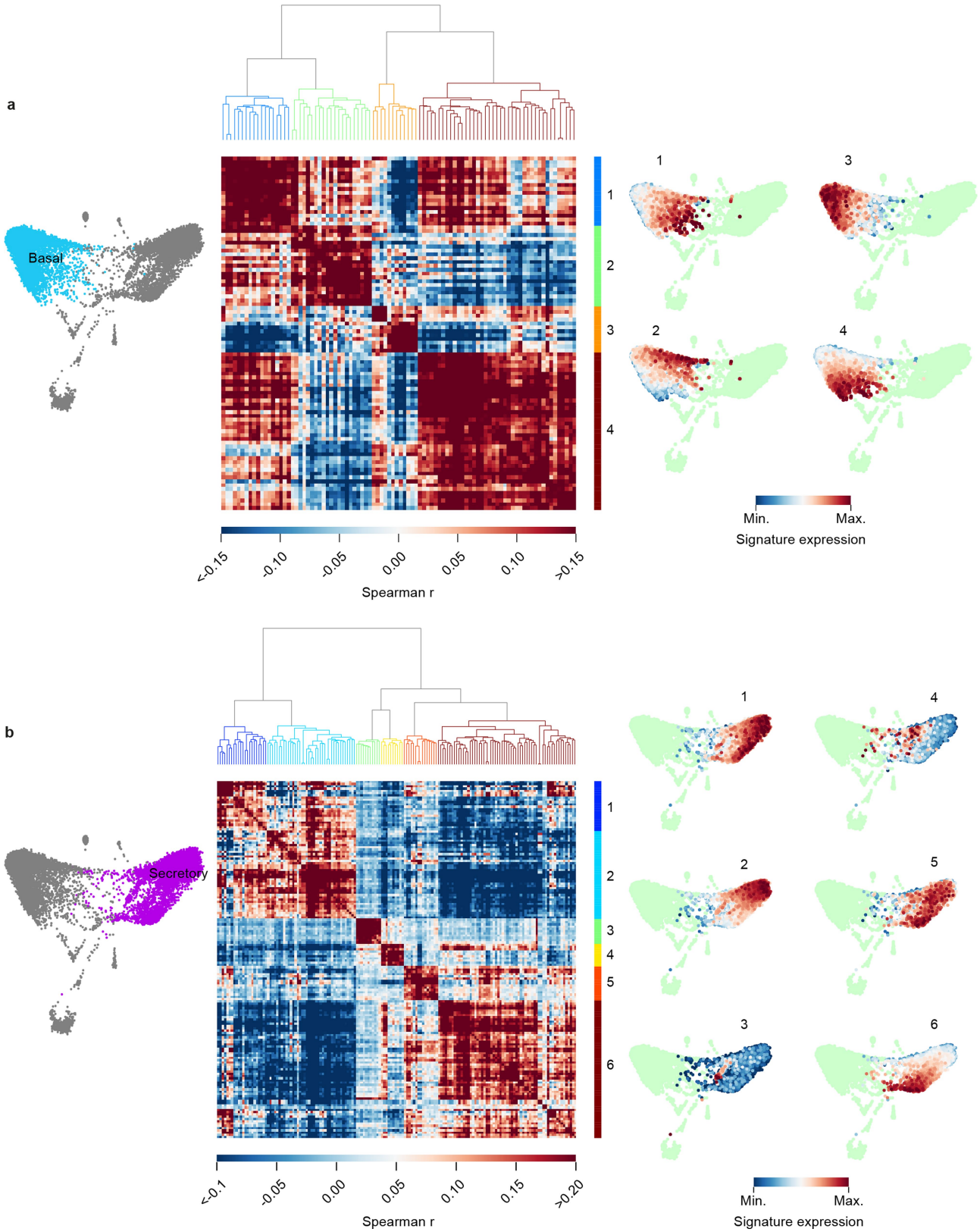
Code availability. Python scripts implementing the methods as described can be obtained upon request.

Data availability. All sequencing data are available in the Gene Ontology Omnibus repository under the accession number GSE102580, the NCBI Sequence Read Archive under the accession number SRR5881096, the Klein laboratory SPRING viewer (https://kleintools.hms.harvard.edu/paper_websites/airway_2018/), and the Single Cell Portal (https://portals.broadinstitute.org/single_cell).

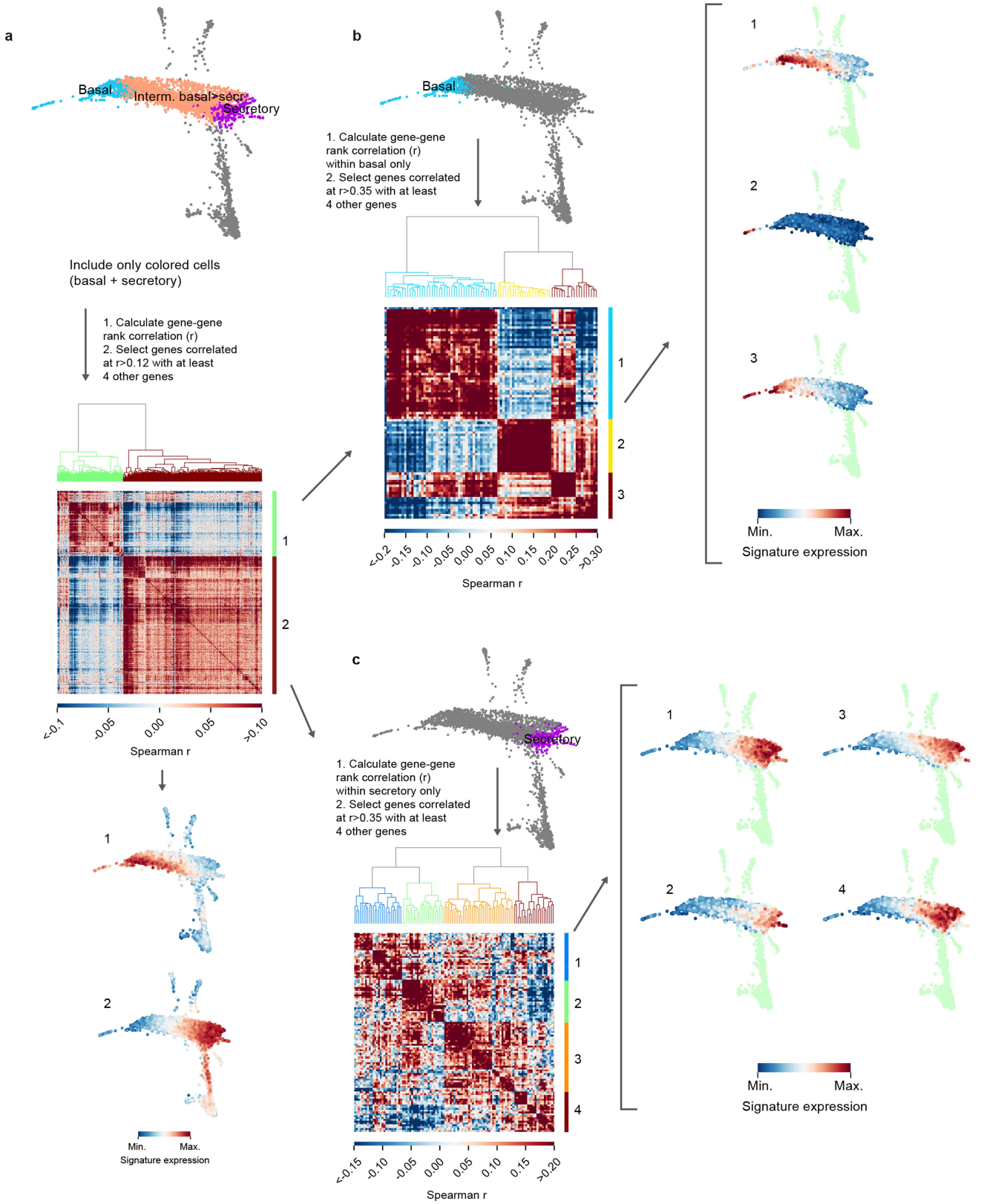
32. Lafkas, D. et al. Therapeutic antibodies reveal Notch control of transdifferentiation in the adult lung. *Nature* **528**, 127–131 (2015).
33. Li, K. et al. Modulation of Notch signaling by antibodies specific for the extracellular negative regulatory region of NOTCH3. *J. Biol. Chem*. **283**, 8046–8054 (2008).
34. Wu, Y. et al. Therapeutic antibody targeting of individual Notch receptors. *Nature* **464**, 1052–1057 (2010).
35. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protocols* **12**, 44–73 (2017).
36. Zhang, H. M. et al. AnimalTFDB 2.0: A resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res*. **43**, D76–D81 (2015).
37. Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. & Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl Acad. Sci. USA* **115**, E2467–E2476 (2018).
38. Tusi, B. K. et al. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018).
39. Borthwick, D. W., Shahbazian, M., Krantz, Q. T., Dorin, J. R. & Randell, S. H. Evidence for stem-cell niches in the tracheal epithelium. *Am. J. Respir. Cell Mol. Biol*. **24**, 662–670 (2001).



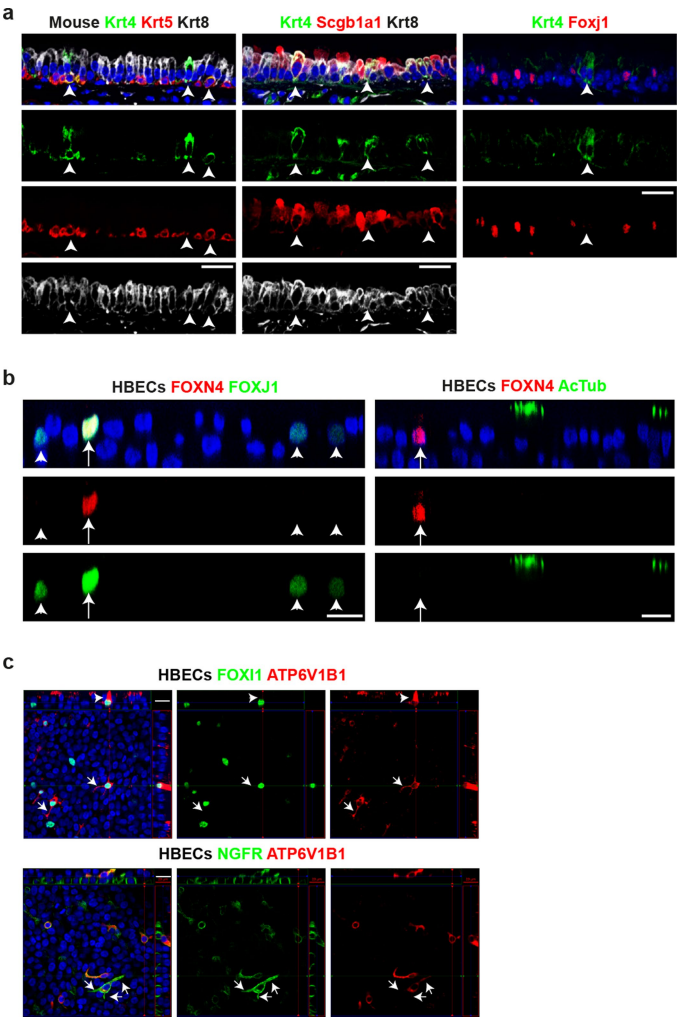
Extended Data Fig. 1 | Atlas of transcription factors, surface molecules and kinases enriched in proximal airway lineages of mouse and human. **a**, **b**, Transcription factors, kinases and surface molecules in mouse (**a**) and human (**b**) identified among the list of cell-type-specific genes that met the following criteria: significantly enriched in lineage (false discovery rate (FDR) <5%, permutation test), expressed at ≥50 transcripts per million (TPM), expressed in marked lineage at least 1.5× higher than second highest cluster and highest in marked lineage for 4/4 (mouse) or 2/3 (human) biological replicates. **c**, Pairwise correlation of cell populations identified by single-cell RNA-seq. Genes used for correlation analysis were the 20% most variable genes (identified as described[8]) of all genes expressed at a level of at least 3 UMI counts in at least 3 cells. Ward's method was used for hierarchical clustering.



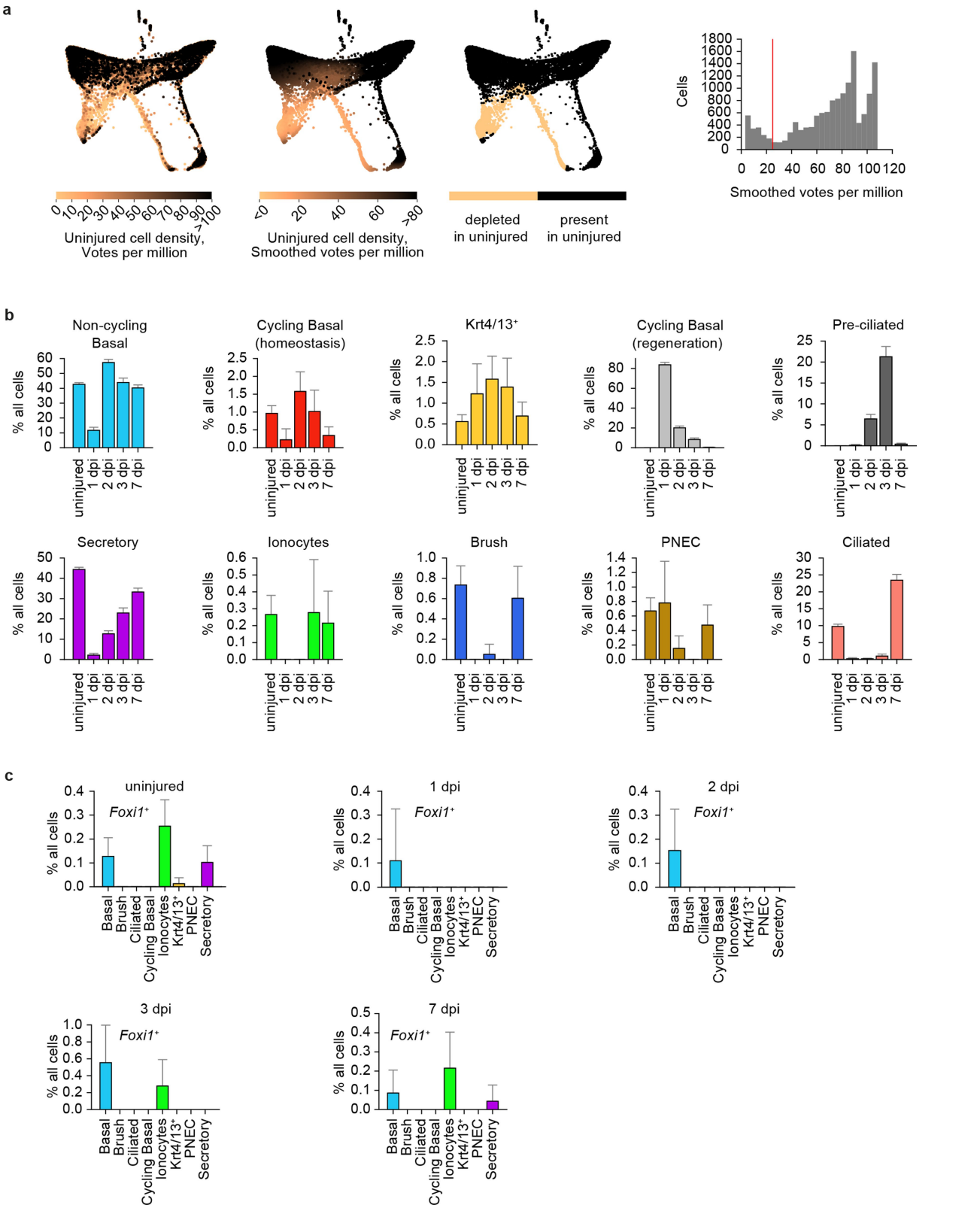
Extended Data Fig. 2 | Gene modules identified in mouse tracheal lineages. Gene modules were identified by selecting variable genes within the given population that were correlated with at least 4 other genes with rank correlation >0.2. **a**, **b**, Gene–gene correlation heat map shows 4 gene modules in mouse airway basal cells (**a**) and 6 gene modules in mouse airway secretory cells (**b**). SPRING plots show where gene modules are expressed in a given population. Multiple genes are combined in a single signature defined as the mean rank of expression (dense ranking).



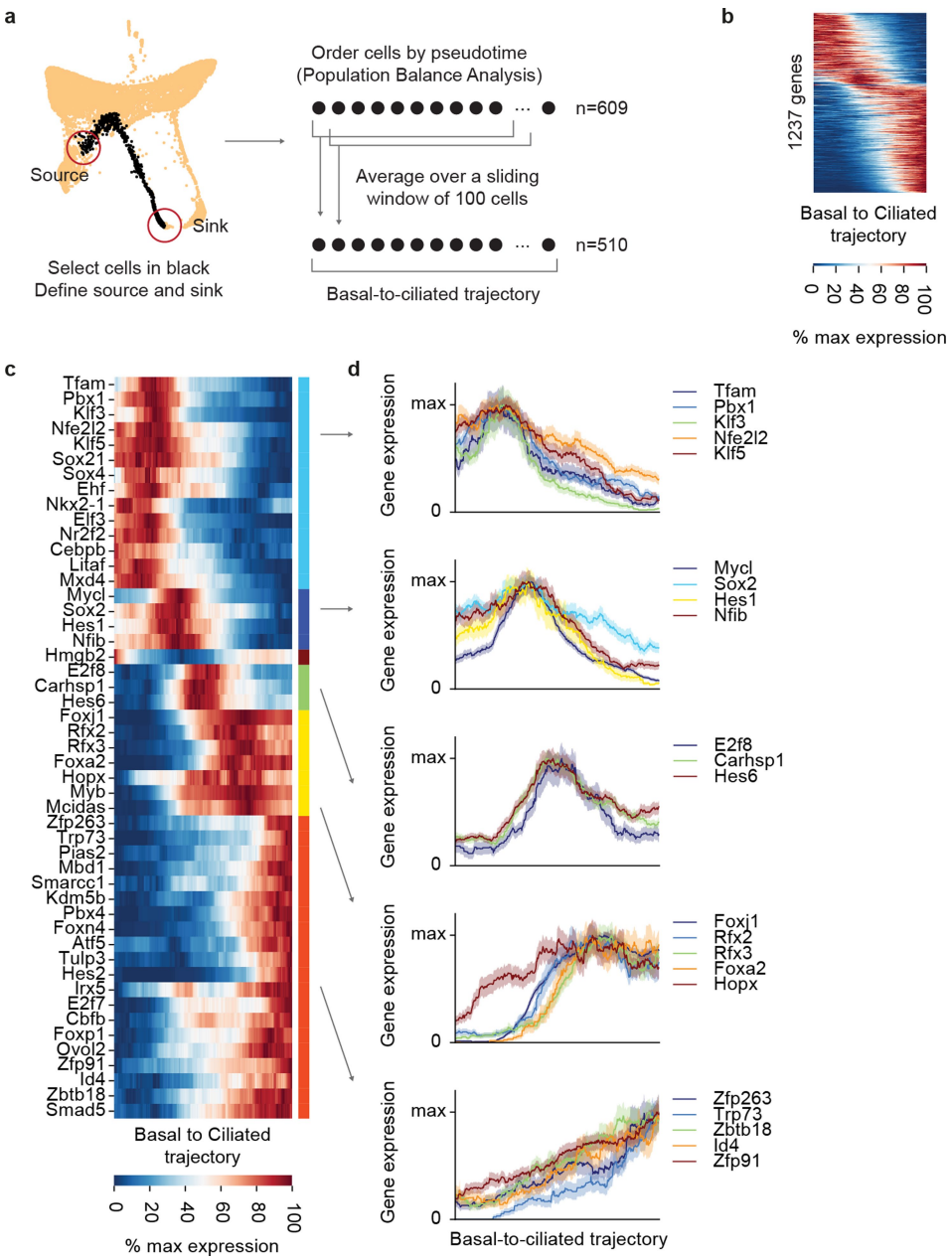
Extended Data Fig. 3 | Gene modules identified in human bronchial lineages. a, Two major modules of anti-correlated genes were identified by selecting variable genes within the basal-to-secretory continuum that were correlated with at least 4 other genes with rank correlation >0.12. Genes within each module were then separately considered within basal and secretory cells, keeping genes with a correlation >0.35 with at least 4 other genes. **b**, **c**, Gene–gene correlation heat map shows 3 gene modules in human airway basal cells (**b**) and 4 gene modules in human airway secretory cells (**c**). SPRING plots show where gene modules are active in a given population. Multiple genes are combined in a single signature defined as the mean rank of expression (dense ranking).



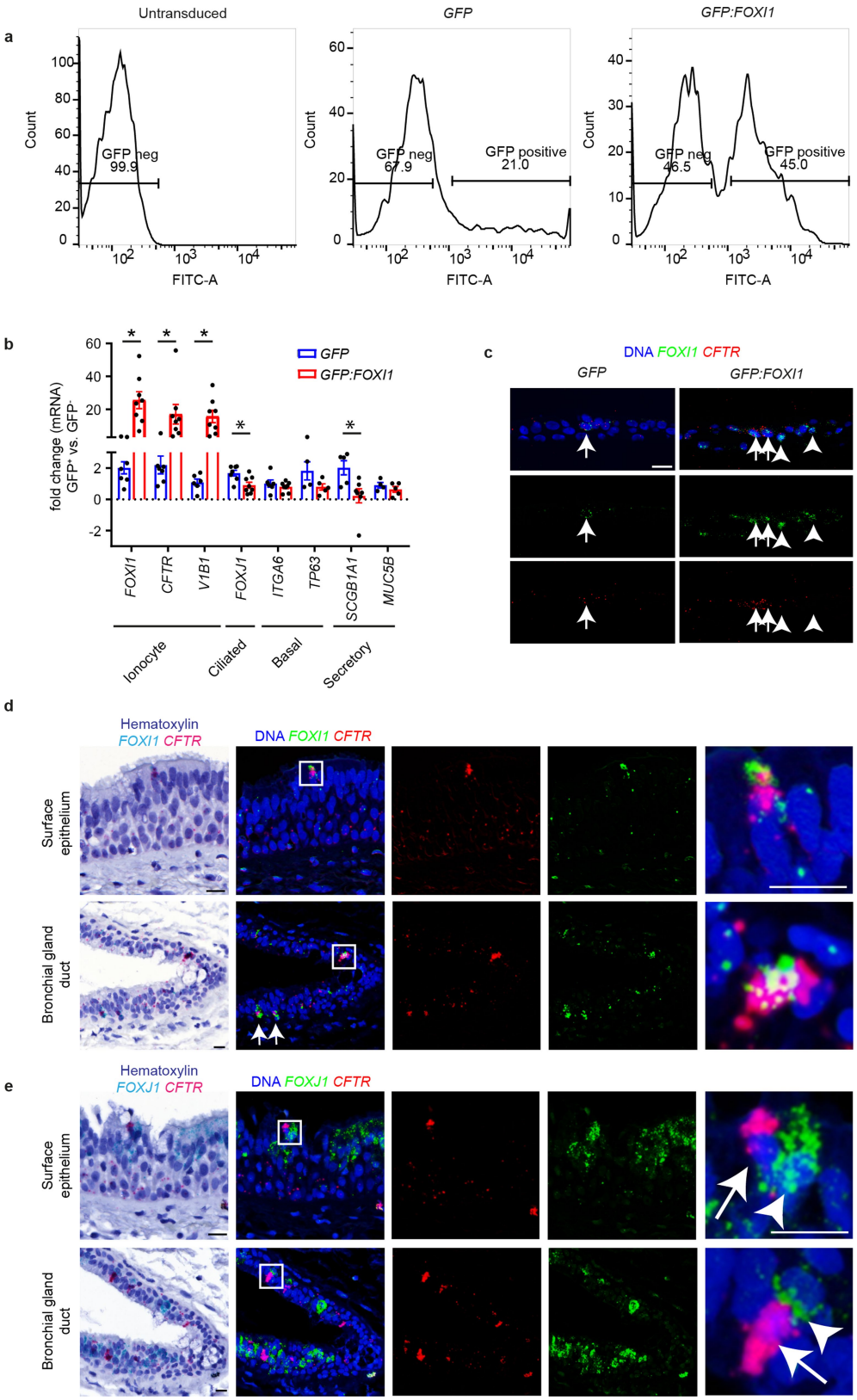
Extended Data Fig. 4 | Validation of novel lineages in mouse and human by immunofluorescence. a, Immunofluorescence in mouse tracheal epithelium for Krt4 (green, arrowheads), Krt5 (basal), Krt8 (luminal), Scgb1a1 (club, secretory) and Foxj1 (ciliated) ($n = 3$ mice). **b**, Immunofluorescence in differentiated HBEC cultures for FOXN4 (red, arrows), FOXJ1 (arrowheads mark FOXJ1^{low} cells) and acetylated α-tubulin (cilia) ($n = 2$ donors). **c**, Immunofluorescence in HBEC cultures for the ionocyte markers FOXI1, ATP6V1B1 and NGFR ($n = 3$ donors). Arrowhead shows apical enrichment of ATP6V1B1. Arrows highlight lateral protrusions. Scale bar, 20 μm.



Extended Data Fig. 5 | Identification of recovery-specific cell states and population dynamics during regeneration. a, Cells from uninjured mouse airway do not equally populate all regions of the SPRING plot of all mouse data combined. Each cell from the uninjured condition voted for its 10 nearest neighbours among all mouse cells profiled, and smoothed vote counts were used as a proxy for uninjured cell density on the map (two left-most plots). By visual inspection of the smooth vote distribution a threshold of 25 votes was chosen to binarize regions of the SPRING plot into present versus depleted in uninjured. **b**, Bar charts representing abundance of rare populations as a fraction of all cells, over time post-injury. Error bars represent the 95% binomial proportion confidence interval (normal approximation). Total number cells = 7,898 from $n = 4$ mice (uninjured), 898 from $n = 1$ mouse (1 dpi), 1,964 from $n = 1$ mouse (2 dpi), 1,082 from $n = 1$ mouse (3 dpi) and 2,321 from $n = 3$ mice (7 dpi). **c**, Bar charts showing the fraction of all cells in each population that express *Foxi1* during recovery. Values shown correspond to the fraction of all cells at each time point (cell and mouse numbers as in **b** above). Error bars defined as in **b**.

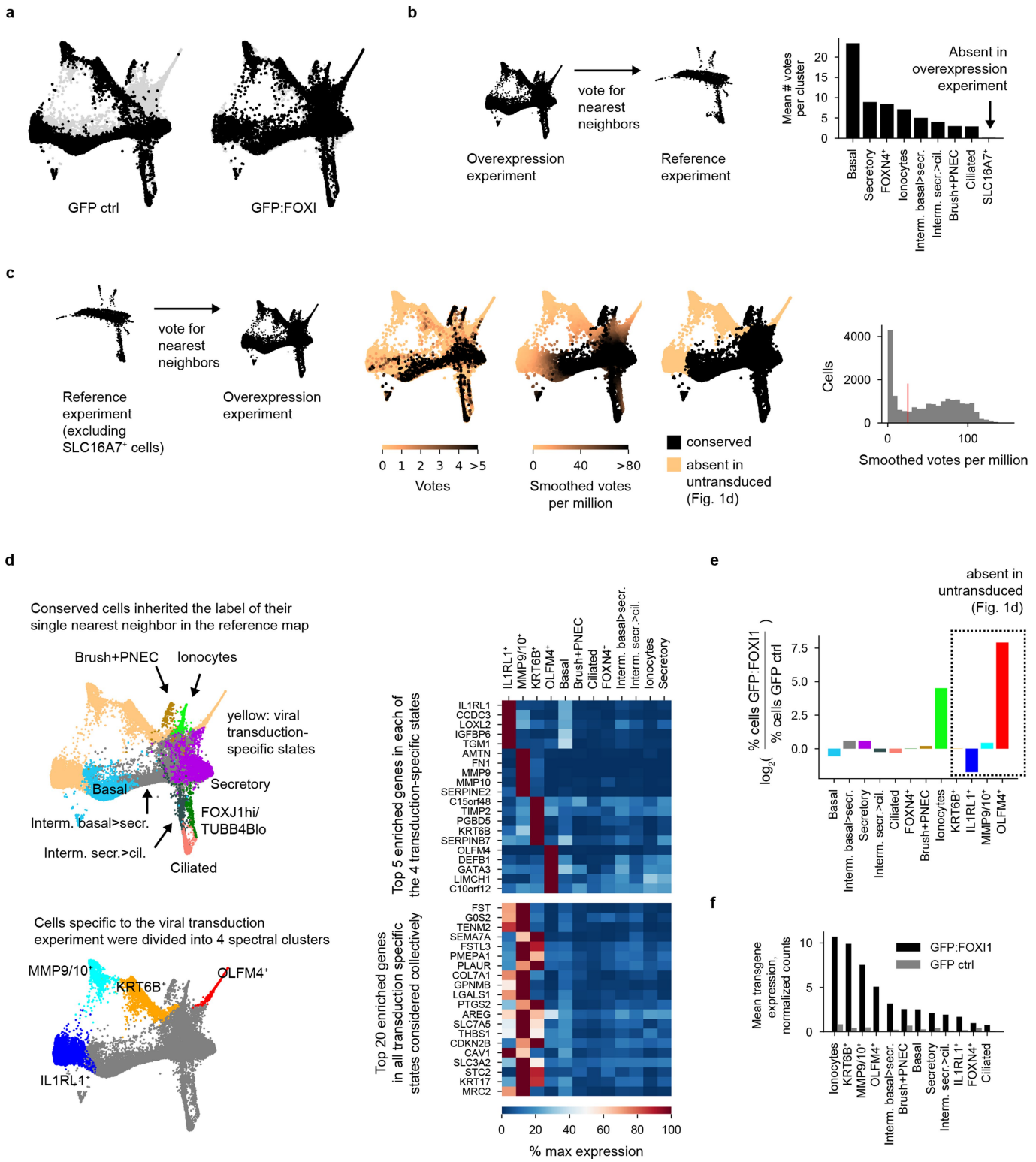


Extended Data Fig. 6 | Analysis of basal-to-ciliated differentiation trajectory following injury. a, Population balance analysis (PBA, see Methods) was used to order 609 cells highlighted in black along the pseudotime of their basal-to-ciliated progression, followed by application of a moving average over a window of 100 cells. The resulting ordering of averaged cells is referred to as the basal-to-ciliated trajectory. PBA requires manually selecting source and sink cells for calculating the pseudotime. **b**, Heat map of the 1,237 differentially expressed genes along the basal-to-ciliated trajectory (permutation test, FDR <5%, fold-change$_{max} \geq 2$; see Methods). Genes ordered by expression-weighted mean position, defined for an expression time series x_t as $\tau = \frac{\sum_t t x_t}{\sum_t x_t}$. **c**, Heat map of transcription factors only. Hierarchical clustering revealed six major clusters of correlated genes. Clusters were ordered by mean expression weighted mean position. **d**, Plots of up to 5 transcription factors sampled from each cluster. The *y* axis shows the average expression of a gene within the window of 100 cells ± s.e.m. (or +1/[window size] for mean values of zero), normalized to the maximum value. The total trajectory includes 609 cells.

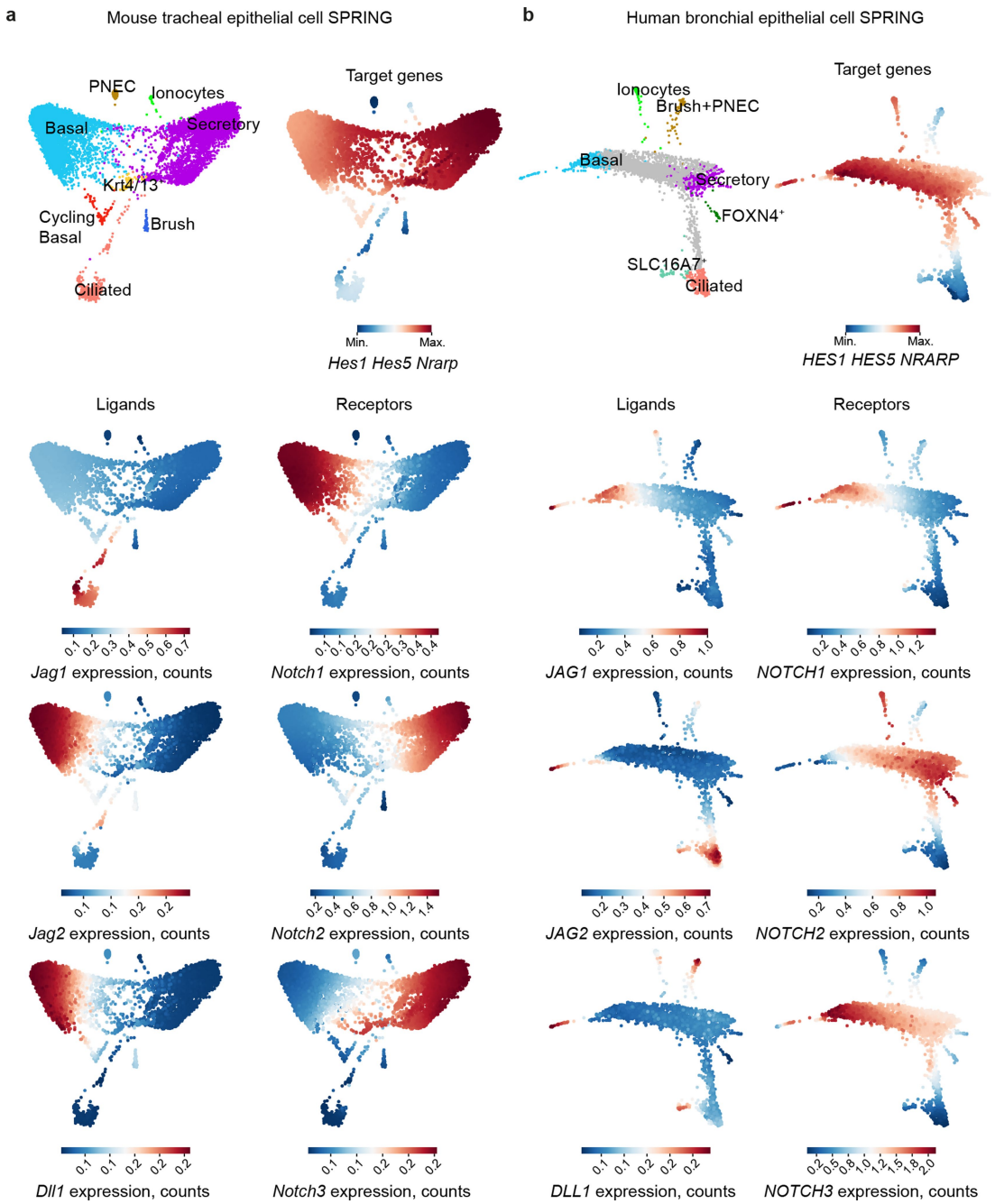


Extended Data Fig. 7 | See next page for caption.

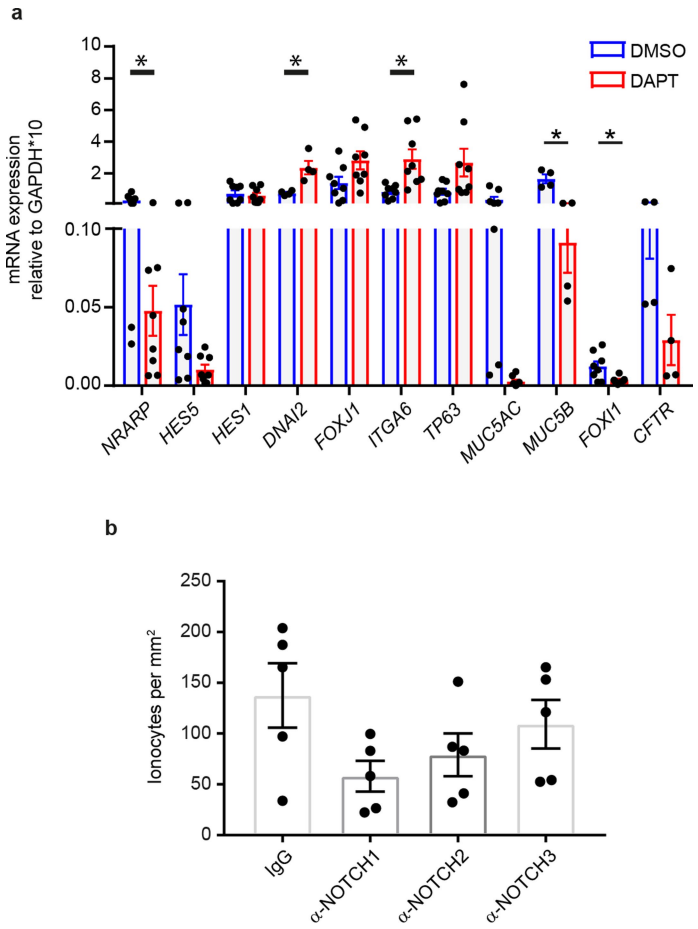
Extended Data Fig. 7 | Specification and characterization of FOXI1 lineage in human bronchial epithelium. a, HBECs were transduced at seeding with *GFP* or *GFP:FOXI1* lentivirus, differentiated and sorted for GFP (shown is representative gating strategy, $n = 12$). **b**, Fold change in transduced cells (GFP^+) compared to non-transduced cells (GFP^-) was determined by RT–qPCR normalized to *GAPDH*. Pooled data from 2 donors transduced with *GFP* ($n = 7$ samples) or *GFP:FOXI1* ($n = 8$ samples) are represented as mean ± s.e.m. *P* value: *FOXI1*, 0.001; *CFTR*, 0.04; *ATP6V1B1*, 0.006; *FOXJ1*, 0.01; *SCGB1A1* , 0.02; two-tailed *t*-test. **c**, Fluorescent in situ hybridization (RNAscope) for *FOXI1* and *CFTR* in HBEC culture transduced with *GFP* or *GFP:FOXI1*. Note that while there is an increase in *FOXI1/CFTR* co-labelled cells, not all *FOXI1* cells express *CFTR*. (arrowheads versus arrows) ($n = 2$ independent experiments in 2 donors). **d**, **e**, Chromogenic in situ hybridization (RNAscope) in primary human bronchial tissue surface epithelium and gland ducts for *CFTR* and *FOXI1* (**d**) or *FOXJ1* (**e**). Chromogenic signals were split and pseudocoloured to reveal individual channels; inset regions are shown at higher magnification on the right. Note that *CFTR* is highly enriched in $FOXI1^+$ but not $FOXJ1^+$ cells ($n = 1$ donor, 5 regions of bronchial tree analysed). Scale bars, 20 μm.



Extended Data Fig. 8 | Single-cell RNA-seq analysis of HBECs transduced with *GFP* and *GFP:FOXI1*. **a**, SPRING plot combining cells transduced with *GFP* ($n = 9{,}436$) or *GFP:FOXI1* ($n = 10{,}330$), with each of the two conditions highlighted in black (total cells $n = 19{,}766$). **b**, The SLC16A7^{+} population was identified to be absent in the viral transduction experiment after mapping single-cell transcriptome onto the reference state map. Each cell from the viral transduction experiment voted for its nearest neighbour in the reference experiment. The bar chart on the right shows the average number of votes per cluster. **c**, Cell states unique to the viral transduction experiment were identified as detailed in Extended Data Fig. 5a. **d**, Cells representing states also found in the reference experiment (conserved cells) inherited the label of their single nearest neighbour in the reference map. Cells specific to the viral transduction experiment were divided into four clusters by spectral clustering, with their top five enriched genes shown in the top part of the heat map (right). Enrichment of gene *g* in population *i* is defined as the fold change in expression of *g* in *i* versus the second highest expresser. A pseudo value of 10 TPM was added before calculating the fold change, and only genes expressed at >50 TPM in at least one cluster were considered. The bottom of the heat map shows the top 20 enriched genes identified treating all four transduction-specific states as one population. **e**, Bar chart showing fold changes in population size following *GFP:FOXI1* versus *GFP* transduction (extension of Fig. 3f). **f**, Expression of transgene in identified cell populations.



Extended Data Fig. 9 | Notch pathway component enrichment in airway lineages. a, **b**, SPRING plots show enrichment of Notch pathway components in mouse (**a**) and human (**b**) airway lineages. Normalized counts are shown for the Notch ligands *JAG1*, *JAG2* and *DLL1*, and the Notch receptors *NOTCH1*, *NOTCH2* and *NOTCH3*. The Notch target gene signature combines *HES1*, *HES5* and *NRARP* into a single gene signature, defined as the mean expression rank (dense ranking). All gene expression and signature values are smoothed (see Methods for smoothing).



Extended Data Fig. 10 | Inhibition of Notch signalling decreases ionocyte markers in HBECs. a, Expression of Notch target genes and airway lineage markers in cultures treated with 3.3 μM DAPT compared to cultures treated with DMSO. Notch target genes (*NRARP* $P = 0.03$, *HES5*) and secretory cell markers (*MUC5B* $P = 0.001$, *MUC5AC*) are decreased whereas ciliated cell markers (*FOXJ1*, *DNAI2* $P = 0.01$) and basal cell markers (ITGA6 $P = 0.006$ and TP63) are increased upon DAPT treatment. Note that ionocyte markers (*FOXI1* $P = 0.02$, *CFTR)* are also decreased upon DAPT treatment. Two-tailed *t*-test; $n = 8$ experiments in 2 donors. **b**, FOXI1 cell counts in HBEC cultures treated with antibodies that neutralize individual Notch receptors ($n = 5$ experiments in 2 donors). All data are mean ± s.e.m.

LETTER

https://doi.org/10.1038/s41586-018-0392-8

Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response

Gang Chen[1], Alexander C. Huang[2], Wei Zhang[1,3], Gao Zhang[4], Min Wu[1], Wei Xu[5], Zili Yu[3], Jiegang Yang[1,3], Beike Wang[1,3], Honghong Sun[6], Houfu Xia[3], Qiwen Man[3], Wenqun Zhong[1,3], Leonardo F. Antelo[5], Bin Wu[1], Xuepeng Xiong[3], Xiaoming Liu[6], Lei Guan[1,7], Ting Li[6,7], Shujing Liu[6], Ruifeng Yang[6], Youtao Lu[1], Liyun Dong[6], Suzanne McGettigan[2], Rajasekharan Somasundaram[4], Ravi Radhakrishnan[8], Gordon Mills[9], Yiling Lu[9], Junhyong Kim[1], Youhai H. Chen[6], Haidong Dong[10], Yifang Zhao[3], Giorgos C. Karakousis[6], Tara C. Mitchell[2,5], Lynn M. Schuchter[2,5], Meenhard Herlyn[4], E. John Wherry[11,12], Xiaowei Xu[6]* & Wei Guo[1]*

Tumour cells evade immune surveillance by upregulating the surface expression of programmed death-ligand 1 (PD-L1), which interacts with programmed death-1 (PD-1) receptor on T cells to elicit the immune checkpoint response[1,2]. Anti-PD-1 antibodies have shown remarkable promise in treating tumours, including metastatic melanoma[2–4]. However, the patient response rate is low[4,5]. A better understanding of PD-L1-mediated immune evasion is needed to predict patient response and improve treatment efficacy. Here we report that metastatic melanomas release extracellular vesicles, mostly in the form of exosomes, that carry PD-L1 on their surface. Stimulation with interferon-γ (IFN-γ) increases the amount of PD-L1 on these vesicles, which suppresses the function of CD8 T cells and facilitates tumour growth. In patients with metastatic melanoma, the level of circulating exosomal PD-L1 positively correlates with that of IFN-γ, and varies during the course of anti-PD-1 therapy. The magnitudes of the increase in circulating exosomal PD-L1 during early stages of treatment, as an indicator of the adaptive response of the tumour cells to T cell reinvigoration, stratifies clinical responders from non-responders. Our study unveils a mechanism by which tumour cells systemically suppress the immune system, and provides a rationale for the application of exosomal PD-L1 as a predictor for anti-PD-1 therapy.

Extracellular vesicles, such as exosomes and microvesicles (also known as shedding vesicles), carry bioactive molecules that influence the extracellular environment and the immune system[6–8]. We purified exosomes from a panel of human primary and metastatic melanoma cell lines by differential centrifugation[9–12], and verified them by transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA) (Fig. 1a, b). Proteins associated with the exosomes were then analysed by reverse phase protein array (RPPA), a large-scale antibody-based quantitative proteomics technology[13]. Analysis by RPPA and western blot revealed the presence of PD-L1 in exosomes, and its level was significantly higher in exosomes derived from metastatic melanoma cells compared to those from primary melanoma cells (Fig. 1c, d, Extended Data Fig. 1a). Iodixanol density gradient centrifugation further confirmed the association of PD-L1 with the exosomes (Extended Data Fig. 1b). PD-L1 was also detected in microvesicles, but at a lower level (Extended Data Fig. 1c–e). PD-L1 was also detected in extracellular vesicles generated from mouse metastatic melanoma B16-F10 cells (Extended Data Fig. 1f).

Tumour cell surface PD-L1 can be upregulated in response to IFN-γ secreted by activated T cells, and PD-L1 binds to PD-1 through its extracellular domain to inactivate T cells[2,3,14]. Using immuno-electron microscopy and enzyme-linked immunosorbent assay (ELISA) (Fig. 1e–g), we found that exosomal PD-L1 has the same membrane topology as cell surface PD-L1, with its extracellular domain exposed on the surface of the exosomes. Exosomal PD-L1 binds PD-1 in a concentration-dependent manner, and this interaction can be disrupted by PD-L1-blocking antibodies (Fig. 1h). Furthermore, the level of exosomal PD-L1 secreted by melanoma cells increased markedly upon IFN-γ treatment (Fig. 1f, g, i), and correspondingly, these exosomes displayed increased binding to PD-1 (Fig. 1h).

Exosomes are generated and released through a defined intracellular trafficking route[7,9,10]. Genetic knockdown of the ESCRT subunit Hrs, which mediates the recognition and sorting of exosomal cargos[15], using short hairpin (sh)RNA led to a decrease in the level of PD-L1 in the exosomes and an increase of PD-L1 in the cell (Extended Data Fig. 1g, h). In addition, PD-L1 co-immunoprecipitated with Hrs from cell lysates (Extended Data Fig. 1i). PD-L1 co-localized with Hrs and CD63, an exosome marker, in melanoma cells (Extended Data Fig. 1j, k). Knockdown of Rab27A, which mediates exosome release[16], also blocked PD-L1 secretion via the exosomes (Extended Data Fig. 1l).

To investigate the secretion of exosomal PD-L1 by melanoma cells in vivo, we established human melanoma xenografts in nude mice. Blood from these mice was collected for exosome purification and subsequent detection of human PD-L1 proteins by ELISA (Fig. 2a). Antibodies against human PD-L1 specifically identified human PD-L1 on the circulating exosomes from mice bearing human melanoma xenografts but not the control mice (Fig. 2b, Extended Data Fig. 2a, b). Moreover, the level of circulating exosomal PD-L1 positively correlated with tumour size (Fig. 2c).

PD-L1 has been found in blood samples derived from melanoma patients[17]. Recent studies suggest the presence of PD-L1 in extracellular vesicles isolated from blood samples of patients with cancer, and the level of PD-L1 correlates with pathological features of these patients[18–20]. We purified extracellular vesicles from the plasma of melanoma patients (Extended Data Fig. 2c–g). The level of PD-L1 on the circulating exosomes was significantly higher in patients with metastatic melanoma than in healthy donors (Fig. 2d, Extended Data Figs. 2f, 3a, 3b), whereas there was no or only marginal difference in the number of circulating exosomes or the total protein level on these exosomes (Extended Data Fig. 3c, d). There was less difference in PD-L1 levels in circulating microvesicles compared to the circulating exosomes (Extended Data Fig. 3e). The data analysis and receiver operating characteristic (ROC) curve show that, among all the parameters tested,

[1]Department of Biology, School of Arts & Sciences, University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Key Laboratory of Oral Biomedicine Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan, China. [4]Molecular and Cellular Oncogenesis Program and Melanoma Research Center, The Wistar Institute, Philadelphia, PA, USA. [5]Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [6]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [7]Ministry of Education Key Laboratory of Biomedical Information Engineering, School of Life Science, Xi'an Jiaotong University, Xi'an, China. [8]Department of Bioengineering, School of Engineering, University of Pennsylvania, Philadelphia, PA, USA. [9]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [10]Department of Immunology, College of Medicine, Mayo Clinic, Rochester, MN, USA. [11]Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [12]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. *e-mail: xug@pennmedicine.upenn.edu; guowei@sas.upenn.edu

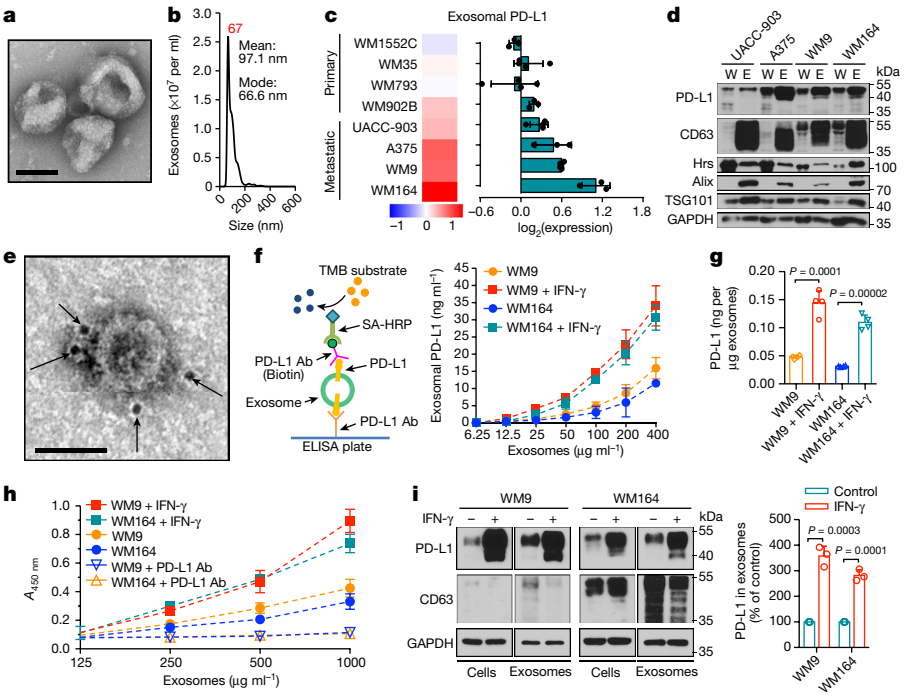


Fig. 1 | Extrafacial expression of PD-L1 on melanoma cell-derived exosomes and its regulation by IFN-γ. a, A representative TEM image of purified exosomes from WM9 cells. Scale bar, 50 nm. **b**, Characterization of purified exosomes using nanoparticle tracking. **c**, RPPA data showing the levels of PD-L1 in exosomes secreted by primary or metastatic melanoma cell lines ($n = 3$ for WM1552C, WM902B, A375, WM164; $n = 4$ for WM35, WM793, UACC-903, WM9). See Extended Data Fig. 1a for statistical analysis. **d**, Immunoblots for PD-L1 in the whole cell lysate (W) and purified exosomes (E) from different metastatic melanoma cell lines. All lanes were loaded with the same amount of total protein. **e**, A representative TEM image of WM9 cell-derived exosomes immunogold-labelled with anti-PD-L1 antibodies. Arrowheads indicate 5-nm gold particles. Scale bar, 50 nm. **f**, Schematic (left) of ELISA to measure PD-L1 concentration (right) on the surface of exosomes isolated from indicated cell types. TMB, 3, 3′, 5, 5′-tetramethylbenzidine; SA-HRP, streptavidin–horseradish peroxidase. **g**, ELISA of PD-L1 on exosomes from melanoma cells, with or without IFN-γ treatment. **h**, PD-l binding of exosomes with IFN-γ or blocking PD-L1 antibody (PD-L1 Ab) (see Methods). **i**, Western blot analysis of PD-L1 in whole cells and exosomes from IFN-γ-treated cells and control cells. All lanes were loaded with the same amount of total protein (left). Quantification of exosomal PD-L1 by western blotting (right). Results shown represent three (**a**, **b**) or two (**d**, **e**) independent experiments. Data are mean ± s.d. of three (**f**, **h**, **i**) or four (**g**) independent biological replicates. *P* values are from a two-sided unpaired *t*-test (**g**, **i**). Full gel source data (**d**, **i**) are shown in Supplementary Fig. 1.

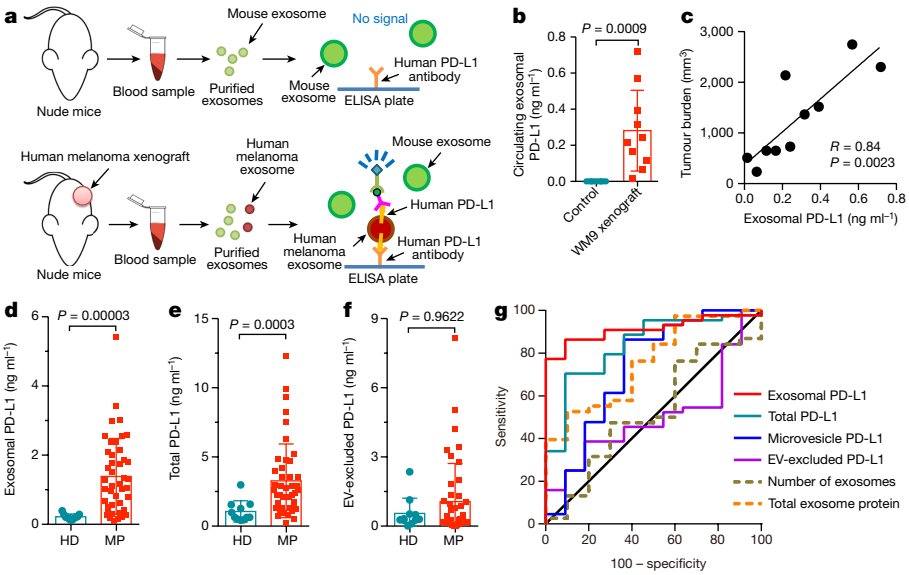


Fig. 2 | The level of PD-L1 on circulating exosomes distinguishes patients with metastatic melanoma from healthy donors. a, ELISA of human PD-L1 on exosomes in plasma samples from mice with human melanoma xenograft. **b**, Levels of PD-L1 on exosomes isolated from the plasma samples of control nude mice or mice bearing human WM9 melanoma xenograft, measured by ELISA ($n = 10$). **c**, Pearson correlation between the exosomal PD-L1 in plasma and tumour burden in xenograft-bearing nude mice ($n = 10$). **d**–**f**, ELISA of circulating exosomal PD-L1 (**d**), total PD-L1 (**e**) or extracellular vesicle (EV)-excluded PD-L1 (**f**) in healthy donors (HD, $n = 11$) and melanoma patients (MP, $n = 44$). The exosomes were purified using the exosome isolation kit. **g**, ROC curve analysis for the indicated parameters in patients with metastatic melanoma compared to healthy donors. Data are mean ± s.d. *P* values are from a two-sided unpaired *t*-test (**b**, **d**–**f**).

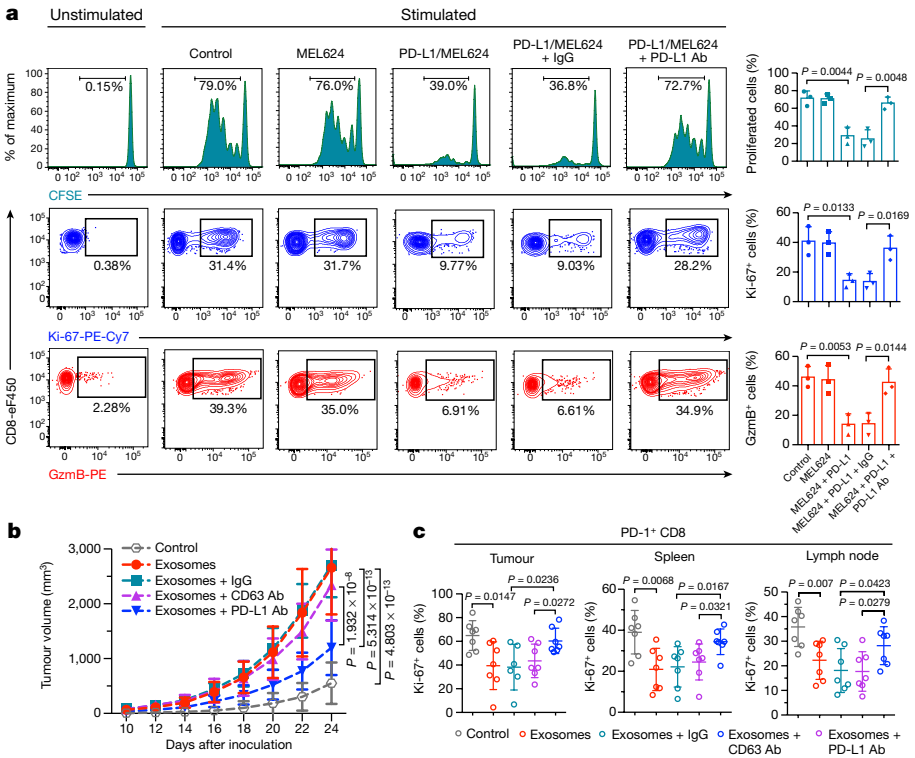


Fig. 3 | Exosomal PD-L1 inhibits CD8 T cells and facilitates the progression of melanoma in vitro and in vivo. a, Representative histogram of CFSE-labelled human peripheral CD8 T cells (top left) and representative contour plots of human peripheral CD8 T cells examined for the expression of Ki-67 (middle left) and granzyme B (GzmB) (bottom left) after indicated treatments. The proportions of cells with diluted CFSE dye, or positive Ki-67 or GzmB expression are shown on the right ($n = 3$ independent biological experiments). **b**, Growth curve of PD-L1(KD) B16-F10 tumours with indicated treatments ($n = 7$ mice per group). **c**, The proportions of Ki-67$^+$PD-1$^+$ CD8 TILs or splenic or lymph node CD8 T cells after indicated treatments ($n = 6$ for tumour samples of the EXO-IgG group, and $n = 7$ for all the other groups). See Extended Data Fig. 8d for representative contour plots. Data are mean ± s.d. (**a**–**c**). *P* values are from a two-sided unpaired *t*-test (**a**, **c**) or two-way ANOVA (**b**).

the level of circulating exosomal PD-L1 best distinguished melanoma patients from healthy donors (Fig. 2d–g, Extended Data Fig. 3e, f).

The current model for PD-L1-mediated immunosuppression is based on the interaction between PD-L1 on the tumour cell surface and PD-1 on CD8 T cells. Here we tested whether exosomal PD-L1 inhibits CD8 T cells. First, we used confocal microscopy to show a physical interaction between tumour exosomes and CD8 T cells purified from human peripheral blood (Extended Data Fig. 4a, b). Flow cytometry analyses further indicated that the level of interaction was higher for activated CD8 T cells than for non-activated counterparts (Extended Data Fig. 4c). Moreover, exosomes derived from melanoma cells treated with IFN-γ exhibited a higher level of binding to CD8 T cells (Extended Data Fig. 4d). Next, we tested the effect of exosomal PD-L1 on CD8 T cells, taking advantage of MEL624 cells, which do not express endogenous PD-L1 (Extended Data Fig. 5a–d) and other immunosuppressive proteins such as FasL and TRAIL[1]. Exosomes derived from MEL624 cells expressing exogenous PD-L1 inhibited the proliferation, cytokine production and cytotoxicity of CD8 T cells, as demonstrated by the decreased proportion of cells containing diluted carboxyfluorescein succinimidyl ester (CFSE, a cell division-tracking dye), reduced expression of Ki-67 and granzyme B (GzmB), and the inhibited production of IFN-γ, IL-2, and TNF (Fig. 3a, Extended Data Fig. 5e, f). Pre-treatment of the exosomes with anti-PD-L1 antibodies nearly abolished these effects. Similar effects were observed using exosomes secreted from WM9 cells, which express endogenous PD-L1 (Extended Data Fig. 5e–h). Exosomes derived from mouse melanoma B16-F10 cells also inhibited the proliferation and cytotoxicity of mouse splenic CD8 T cells (Extended Data Fig. 6a–d). Pre-treating OT-I T cells (which specifically recognize OVA peptide) (Extended Data Fig. 6e) with B16-F10 cell-derived exosomes inhibited their ability to kill their target cells (Extended Data Fig. 6f). Extracellular vesicles from human lung and breast cancer cells also contain immunosuppressive PD-L1, mostly of which is in exosomes, and PD-L1 expression is also upregulated by IFN-γ in some of these cell lines (Extended Data Fig. 7a–e).

To examine the effects of exosomal PD-L1 in vivo, we established a syngeneic mouse melanoma model in C57BL/6 mice using B16-F10 cells in which PD-L1 expression had been knocked down (PD-L1(KD) B16-F10) (Extended Data Fig. 8a). Injection of exosomes derived from parental B16-F10 cells promoted the growth of tumours derived from PD-L1(KD) B16-F10 cells, whereas pre-treatment of the exosomes with anti-PD-L1 antibodies, but not with IgG isotype or CD63-blocking antibodies, inhibited the effect (Fig. 3b, Extended Data Fig. 8b, c). The number of tumour-infiltrating CD8 T lymphocytes (TILs) decreased significantly after the injection of exosomes (Fig. 3c, Extended Data Fig. 8d, e). B16-F10 exosomes also decreased the proportion of proliferating PD-1$^+$ CD8 T cells in both spleen and lymph nodes (Fig. 3c, Extended Data Fig. 8d), suggesting that exosomal PD-L1 suppresses anti-tumour immunity systemically.

We then examined the level of PD-L1 on circulating extracellular vesicles in melanoma patients during anti-PD-1 therapy. The pre-treatment level of circulating exosomal PD-L1 was significantly higher in patients who failed to respond to the anti-PD-1 treatment with pembrolizumab (Fig. 4a). The difference was, however, not significant for total circulating PD-L1, and undetectable for PD-L1 on circulating microvesicles, or extracellular vesicle-excluded PD-L1 (Fig. 4b–d). A higher level of circulating exosomal PD-L1 before the treatment was associated with poorer clinical outcomes (Fig. 4e). IFN-γ upregulates exosomal PD-L1 and the pre-treatment levels of IFN-γ were significantly higher in patients who did not respond to pembrolizumab[21]. The level of circulating exosomal PD-L1 positively correlated with the level of circulating IFN-γ and overall tumour burden (Fig. 4f, g), which were shown to be indicative of poor prognosis[21].

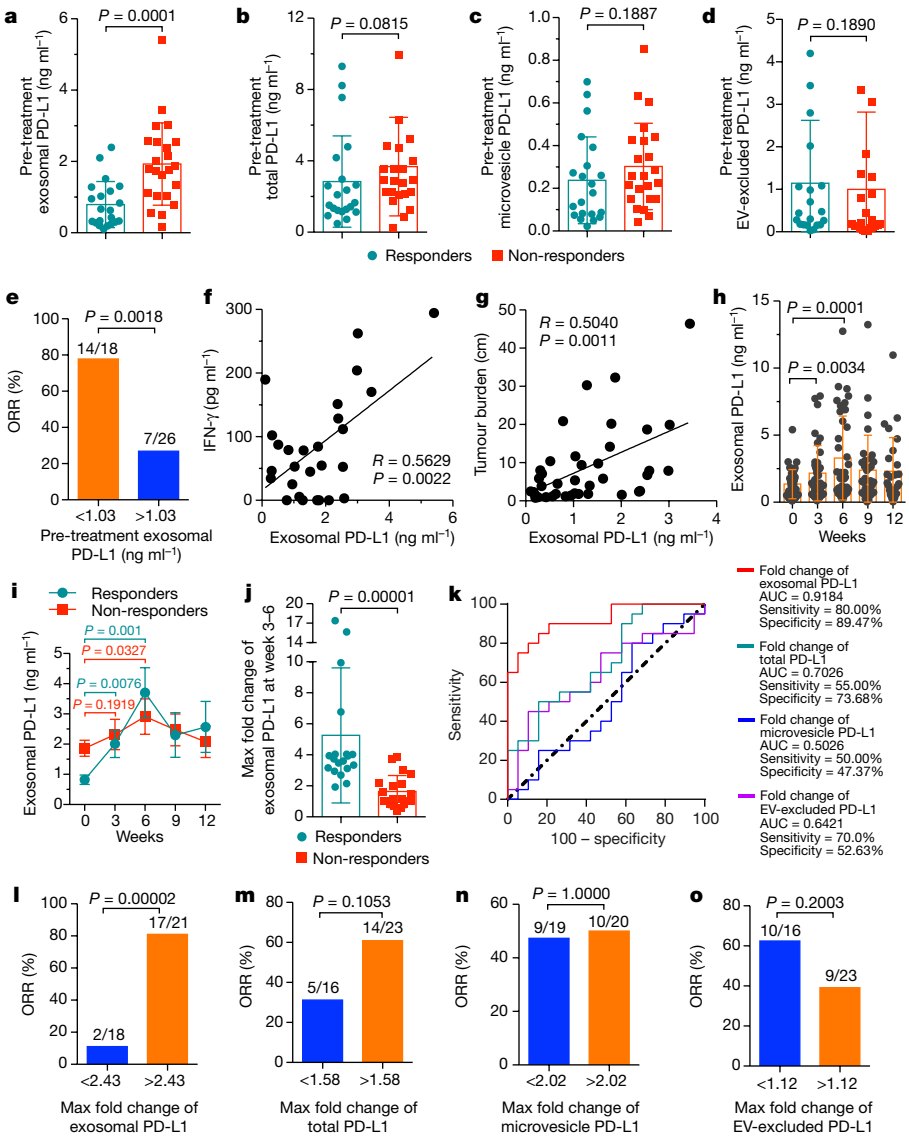


Fig. 4 | The level of circulating exosomal PD-L1 stratifies clinical responders to pembrolizumab and non-responders. **a**–**d**, Comparison of the pre-treatment levels of circulating exosomal PD-L1 (**a**), total PD-L1 (**b**), microvesicle PD-L1 (**c**), or extracellular vesicle-excluded PD-L1 (**d**) between melanoma patients with or without clinical response to pembrolizumab. R, responders; $n = 21$; NR, non-responders; $n = 23$. **e**, Objective response rate (ORR) for patients with high and low pre-treatment levels of circulating exosomal PD-L1. **f**, **g**, Pearson correlation of the IFN-γ level (**f**, $n = 27$) or overall tumour burden (**g**, $n = 39$) to the exosomal PD-L1 level in the plasma of patients with melanoma. **h**, Circulating exosomal PD-L1 at serial time points pre-treatment and on-treatment ($n = 39$). **i**, Circulating exosomal PD-L1 in clinical responders ($n = 19$) and non-responders ($n = 20$) at serial time points pre- and on-treatment. **j**, Comparison of the maximum fold change of circulating exosomal PD-L1 at week 3–6 between clinical responders and non-responders. **k**, ROC curve analysis for the maximum fold change of circulating exosomal PD-L1 at week 3–6 in clinical responders compared to non-responders. AUC, area under curve. **l**–**o**, ORR for patients with high and low fold changes of circulating exosomal PD-L1 (**l**), total PD-L1 (**m**), microvesicle PD-L1 (**n**), or extracellular vesicle-excluded PD-L1 (**o**), at weeks 3–6 of treatment. Data are mean ± s.d. $^*P < 0.05$, two-sided unpaired *t*-test (**a**–**d**, **j**), two-sided paired *t*-test (**h**, **i**), or two-sided Fisher's exact test (**e**, **l**–**o**).

Next, we examined the level of circulating exosomal PD-L1 in patients undergoing pembrolizumab therapy. In clinical responders, there were increased levels of PD-L1 on circulating exosomes, mostly within 6 weeks of therapy (Fig. 4h, i). The level of PD-L1 on microvesicles also increased in the same cohort of patients, but to a lesser extent in comparison to exosomes (Extended Data Fig. 9a). Proliferation and reinvigoration of CD8 T cells peaked at week 3 of treatment and preceded the peaking of exosomal PD-L1 at week 6 (Extended Data Fig. 9b). Moreover, in pembrolizumab-responsive patients, both the absolute value and maximal fold change of Ki-67 in PD-1$^+$ CD8 T cells after 3–6 weeks of treatment positively correlated with those of circulating exosomal PD-L1 (Extended Data Fig. 9c, d). The responders displayed a larger increase in the level of circulating exosomal PD-L1 as early as 3–6 weeks following the initial treatment (Fig. 4j). ROC analysis determined that a fold change of 2.43 in exosomal PD-L1 at week 3–6 stratified patients by clinical response to pembrolizumab (Fig. 4k); a fold change in circulating exosomal PD-L1 greater than 2.43 at week 3–6 was associated with a better response to anti-PD-1 therapy by objective response rate (ORR), progression-free and overall survival (Fig. 4l, Extended Data Fig. 9e). The fold increase of total circulating PD-L1, microvesicle PD-L1, and extracellular vesicle-excluded PD-L1 was inferior to that of exosomal PD-L1 for distinguishing responders from non-responders (Fig. 4k, m–o, Extended Data Fig. 9f–h).

Our studies suggest that melanoma cells release PD-L1-positive extracellular vesicles into the tumour microenvironment and circulation to counter the anti-tumour immunity systemically. Since exosomal PD-L1-mediated T cell inhibition can be blocked by antibodies against either PD-L1 or PD-1, our results raise the possibility that disrupting the interaction between exosomal PD-L1 and PD-1 on T cells is a previously unrecognized mechanism in PD-L1/PD-1 blockade-based therapies. The level of PD-L1 on extracellular vesicles is upregulated by IFN-γ, and PD-L1 on extracellular vesicles primarily targets PD-1$^+$ CD8 T cells, which represent the antigen-experienced T cells that secrete IFN-γ. Exosomal PD-L1 may therefore reflect the dynamic interplay between tumour and immune cells. Besides PD-L1, other extracellular vesicle proteins such as FasL may also contribute to immunosuppressive effects[19,22–24]. However, PD-L1 enables exosomes to target predominantly PD-1$^+$ CD8 T cells, allowing tumour cells to counteract the immune pressure at the effector stage. In addition to the interaction between exosomal-PD-L1 and PD-1, the involvement of other molecules including B7 and CD28[25,26] in this process also warrant investigation.

Our study suggests that circulating exosomal PD-L1 prior to and during pembrolizumab treatment may reflect distinct states of anti-tumour immunity. The pre-treatment PD-L1 level may correlate with a role of exosomal PD-L1 in immune dysfunction. High levels of exosomal PD-L1 may reflect the 'exhaustion' of T cells to a stage at which they can no longer be reinvigorated by anti-PD-1 treatment. In on-treatment patients, however, an increase in the level of exosomal PD-L1, following and correlating positively with T cell reinvigoration,

would reflect the presence of a successful anti-tumour immunity elicited by the anti-PD-1 therapy. Although the increase in exosomal PD-L1 in response to IFN-γ could enable tumour cells to adaptively inactivate CD8 T cells, this is futile because the interaction between PD-L1 and PD-1 is blocked by pembrolizumab. We observed no marked increase in exosomal PD-L1 in non-responders. This could be a result of failure to elicit an adequate T cell response, or a resistance mechanism to IFN-γ from tumours. Tumour cells in non-responders may have adaptively downregulated their response to IFN-γ to avoid the detrimental increase in antigen presentation and to escape the anti-proliferative effects induced by IFN-γ[5,27].

Our study offers a rationale for developing circulating exosomal PD-L1 as a predictor for the clinical outcomes of anti-PD-1 therapy, and sheds light on possible causes for the failure of anti-PD-1 therapies experienced by many patients (Extended Data Fig. 10). Tumour PD-L1 has been used as a predictive biomarker for clinical responses to anti-PD-1 therapy[28–30]. Considering the heterogeneity and dynamic changes of PD-L1 expression in tumours, and the invasive nature of tumour biopsy, developing exosomal PD-L1 as a blood-based biomarker could be an attractive option.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0392-8.

Received: 16 October 2017; Accepted: 13 June 2018;
Published online 8 August 2018.

1. Dong, H. et al. Tumor-associated B7–H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat. Med.* **8**, 793–800 (2002).
2. Chen, L. & Han, X. Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future. *J. Clin. Invest.* **125**, 3384–3391 (2015).
3. Topalian, S. L., Taube, J. M., Anders, R. A. & Pardoll, D. M. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat. Rev. Cancer* **16**, 275–287 (2016).
4. Ribas, A. et al. Association of pembrolizumab with tumor response and survival among patients with advanced melanoma. *J. Am. Med. Assoc.* **315**, 1600–1609 (2016).
5. Zaretsky, J. M. et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).
6. Becker, A. et al. Extracellular vesicles in cancer: cell-to-cell mediators of metastasis. *Cancer Cell* **30**, 836–848 (2016).
7. Kalluri, R. The biology and function of exosomes in cancer. *J. Clin. Invest.* **126**, 1208–1215 (2016).
8. Robbins, P. D. & Morelli, A. E. Regulation of immune responses by extracellular vesicles. *Nat. Rev. Immunol.* **14**, 195–208 (2014).
9. Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr. Protoc. Cell Biol.* **30**, 3.22.1–3.22.29 (2006).
10. Colombo, M., Raposo, G. & Thery, C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu. Rev. Cell Dev. Biol.* **30**, 255–289 (2014).
11. Peinado, H. et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat. Med.* **18**, 883–891 (2012).
12. Melo, S. A. et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* **523**, 177–182 (2015).
13. Tibes, R. et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol. Cancer Ther.* **5**, 2512–2521 (2006).
14. Garcia-Diaz, A. et al. Interferon receptor signaling pathways regulating PD-L1 and PD-L2 expression. *Cell Reports* **19**, 1189–1201 (2017).
15. Schmidt, O. & Teis, D. The ESCRT machinery. *Curr. Biol.* **22**, R116–R120 (2012).
16. Ostrowski, M. et al. Rab27a and Rab27b control different steps of the exosome secretion pathway. *Nat. Cell Biol.* **12**, 19–30 (2010).
17. Zhou, J. et al. Soluble PD-L1 as a biomarker in malignant melanoma treated with checkpoint blockade. *Cancer Immunol. Res.* **5**, 480–492 (2017).
18. Theodoraki, M. N., Yerneni, S. S., Hoffmann, T. K., Gooding, W. E. & Whiteside, T. L. Clinical significance of PD-L1$^+$ exosomes in plasma of head and neck cancer patients. *Clin. Cancer Res.* **24**, 896–905 (2018).
19. Ludwig, S. et al. Suppression of lymphocyte functions by plasma exosomes correlates with disease activity in patients with head and neck cancer. *Clin. Cancer Res.* **23**, 4843–4854 (2017).
20. Ricklefs, F. L. et al. Immune evasion mediated by PD-L1 on glioblastoma-derived extracellular vesicles. *Sci. Adv.* **4**, eaar2766 (2018).
21. Huang, A. C. et al. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* **545**, 60–65 (2017).
22. Kim, J. W. et al. Fas ligand-positive membranous vesicles isolated from sera of patients with oral cancer induce apoptosis of activated T lymphocytes. *Clin. Cancer Res.* **11**, 1010–1020 (2005).
23. Strauss, L., Bergmann, C. & Whiteside, T. L. Human circulating CD4$^+$CD25^{high}Foxp3$^+$ regulatory T cells kill autologous CD8$^+$ but not CD4$^+$ responder cells by Fas-mediated apoptosis. *J. Immunol.* **182**, 1469–1480 (2009).
24. Wieckowski, E. U. et al. Tumor-derived microvesicles promote regulatory T cell expansion and induce apoptosis in tumor-reactive activated CD8$^+$ T lymphocytes. *J. Immunol.* **183**, 3720–3730 (2009).
25. Kamphorst, A. O. et al. Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent. *Science* **355**, 1423–1427 (2017).
26. Hui, E. et al. T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition. *Science* **355**, 1428–1433 (2017).
27. Shin, D. S. et al. Primary Resistance to PD-1 Blockade Mediated by JAK1/2 Mutations. *Cancer Discov.* **7**, 188–201 (2017).
28. Reck, M. et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).
29. Patel, S. P. & Kurzrock, R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol. Cancer Ther.* **14**, 847–856 (2015).
30. Tumeh, P. C. et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).

Acknowledgements We thank S. Fuchs and J. Ridley (University of Pennsylvania) for their helpful comments. This work is supported by NIH grants GM111128 and GM085146 to W.G., AI105343, AI108545, AI082630, AI117950, Parker Institute for Cancer Immunotherapy to E.J.W., 2T32CA009615-26 to A.C.H., American Heart Association to G.C., CA114046, CA025874, CA010815, CA193417, CA047159, P50 CA174523 (SPORE) and the Tara Miller Foundation to M.H., L.M.S., G.C.K., T.C.M., W.G., X.Xu., the University of Pennsylvania and the Wistar Institute, the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation to M.H., the CAST foundation 2016QNRC001, 2015QNRC001 to Wuhan University, and the NSFC foundation 81570994 to Y.Z.

Reviewer information *Nature* thanks H. Peinado, T. Tueting and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions G.C. and W.G. conceived the project and designed the experiments. G.C., W.Zha., M.W., Z.Y., J.Y., H.X., Q.M., W.Zho. and X.Xi. purified and characterized extracellular vesicles. G.C., W.Zha., J.Y., B.Wa., W.Zho., X.L., S.L. and L.D. performed the mice experiments. G.C., W.Zha., M.W., J.Y., B.Wa. and B.Wu performed the immunoprecipitation and western blot analysis. G.C., A.C.H., M.W., J.Y., H.S. and R.Y. performed the flow cytometry experiments. G.C., W.Zha. and J.Y. performed the nanoparticle tracking, ELISAs, qPCR, PD-1 binding and immunofluorescence staining experiments. G.C., Z.Y., H.X. and Q.M. performed the electron microscopy experiments. G.C., W.Zha., H.S. and J.Y. performed the T cell proliferation and tumour-killing experiments. G.C., H.S., L.G. and T.L. generated the bone marrow-derived dendritic cells and performed the antigen cross-presentation experiments. G.M. and Yi.L. performed the RPPA experiments. G.C.K. and X.Xu. performed pathological analyses. W.X., L.F.A., S.M., R.S., T.C.M. and L.M.S. provided human samples and associated clinical data. H.D. and M.H. provided melanoma cell lines. G.C., A.C.H., W.Zha., G.Z., M.W., W.X., Z.Y., Yo.L., R.R., G.M., J.K., Y.H.C., H.D., Y.Z., T.C.M., L.M.S., E.J.W., X.Xu. and W.G. analysed and interpreted the data. G.C. and W.G. wrote the paper. G.Z., A.C.H., Y.H.C., H.D., M.H., E.J.W. and X.Xu. edited the paper. All authors have read and approved the final manuscript.

Competing interests W.G., G.C. and X.Xu. are listed as inventors on a patent owned by the University of Pennsylvania related to this work. W.G. and X.Xu. serve on the Scientific Advisory Board and have equities in Exo Bio, a company that has licensed the patent from the University of Pennsylvania.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0392-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0392-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to X.Xu. or W.G.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Cell culture. The A375 human melanoma and B16-F10 mouse melanoma cells were purchased from ATCC. The control and PD-L1-overexpressing human melanoma MEL624 cells were provided by H. Dong (Mayo Clinic). Mouse melanoma B16 cells stably expressing chicken OVA (B16-OVA) were provided by H. C. J. Ertl (The Wistar Institute). The UACC-903 human melanoma cells were provided by M. Powell (Stanford University). The melanoma cell lines WM1552C, WM35, WM793, WM902B, WM9 and WM164 presented in this study were established in M. Herlyn's laboratory (The Wistar Institute). All cell lines were authenticated by DNA fingerprinting, and were tested routinely before use to avoid mycoplasma contamination. Human melanoma cell lines MEL624, PD-L1/MEL624, WM1552C, WM35, WM902B, WM793, UACC-903, WM9, A375 and WM164 were cultured in RPMI 1640 medium (Invitrogen) supplemented with 10% (v/v) fetal bovine serum (FBS) (Invitrogen). B16-F10 and B16-OVA cells were cultured in DMEM (Sigma) supplemented with 10% (v/v) FBS. For stimulation with IFN-γ, cells were incubated with 100 ng/ml of recombinant human or mouse IFN-γ (Peprotech) for 48 h.

Generation of stable Hrs, Rab27a or PD-L1 knockdown melanoma cells. Short hairpin RNAs (shRNAs) against human *Hrs* (also known as *HGS*) (NM_004712, GCACGTCTTTCCAGAATTCAA, GCATGAAGAGTAACCACAGC), human *RAB27A* (NM_004850, GCTGCCAATGGGACAAACATA, CAGGAGAGGTTT CGTAGCTA) (gift from A. Weaver, Vanderbilt University), mouse *PD-L1* (also known as *Cd274*) (NM_021893, GCGTTGAAGATACAAGCTCAA) or scrambled shRNA-control (Addgene) were packaged into lentiviral particles using 293T cells co-transfected with the viral packaging plasmids. Lentiviral supernatants were harvested 48–72 h after transfection. Cells were infected with filtered lentivirus and selected by 2 μg/ml puromycin.

Patients and specimen collection. Patients with stage III to IV melanoma (Supplementary Table 1) were enrolled for treatment with pembrolizumab (2 mg/kg by infusion every 3 weeks) under an Expanded Access Program at Penn (http://clinicaltrials.gov identifier NCT02083484) or with commercial Keytruda. Patients gave consent in writing for blood collection under the University of Pennsylvania Abramson Cancer Center's melanoma research program tissue collection protocol UPCC 08607 in accordance with the ethics committee and The Institutional Review Board of the University of Pennsylvania. Peripheral blood was obtained in sodium heparin tubes before each pembrolizumab infusion every 3 weeks for 12 weeks. Clinical response was determined as best response based on immune-related RECIST (irRECIST) using unidimensional measurements[31]. The assessment of clinical responses for patients was performed independently in a double-blind fashion. Blood samples from healthy donors were collected at The Wistar Institute after approval by the ethics committee and Institutional Review Board of The Wistar Institute. Written consent was obtained from each healthy donor before blood collection. All experiments involving blood samples from healthy donors were performed in accordance with relevant ethical regulations.

Flow cytometry of patients' PBMCs. Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll gradient and stored using standard protocols. Cryopreserved PBMC samples from pretreatment, cycles 1–4 (weeks 3–12) were thawed and analysed by flow cytometry as previously described[21]. In brief, live or dead cell discrimination was performed using Live/Dead Fixable Aqua Dead Cell Stain Kit (Life Technologies). Cell surface staining was performed for 30 min at 4 °C. Intracellular staining was performed for 60 min on ice after using a fixation/permeabilization kit (eBioscience).

Purification of extracellular vesicles. For exosome purification from cell culture supernatants, cells were cultured in media supplemented with 10% exosome-depleted FBS. Bovine exosomes were depleted by overnight centrifugation at 100,000*g*. Supernatants were collected from 48–72 h cell cultures and extracellular vesicles were purified by a standard differential centrifugation protocol[9–12]. In brief, culture supernatants were centrifuged at 2,000*g* for 20 min to remove cell debris and dead cells (Beckman Coulter, Allegra X-14R). Microvesicles were pelleted after centrifugation at 16,500*g* for 45 min (Beckman Coulter, J2-HS) and resuspended in PBS. Supernatants were then centrifuged at 100,000*g* for 2 h at 4 °C (Beckman Coulter, Optima XPN-100). The pelleted exosomes were suspended in PBS and collected by ultracentrifugation at 100,000*g* for 2 h.

For purification of circulating extracellular vesicles by differential centrifugation, venous citrated blood from melanoma patients or healthy donors was centrifuged at 1,550*g* for 30 min to obtain cell-free plasma (Beckman Coulter, Allegra X-14R). Then, 1 ml of the obtained plasma was centrifuged at 16,500*g* for 45 min (Eppendorf, 5418R). The pelleted microvesicles were suspended in PBS. The collected supernatants were then centrifuged at 100,000*g* for 2 h at 4 °C (Beckman Coulter, Optima™ MAX-XP) to pellet the exosomes. For purification of circulating exosomes using the exosome isolation kit, cell-free plasma was first centrifuged at 16,500*g* for 45 min (Eppendorf, 5418R) to pellet large membrane vesicles. Exosomes were then purified from the supernatants using the exosome isolation kit (Invitrogen, Cat# 4484450).

Characterization of purified exosomes. For verification of purified exosomes using electron microscopy, purified exosomes suspended in PBS were dropped on formvar carbon-coated nickel grids. After staining with 2% uranyl acetate, grids were air-dried and visualized using a JEM-1011 transmission electron microscope. For immunogold labelling, purified exosomes suspended in PBS were placed on formvar carbon-coated nickel grids, blocked, and incubated with mouse anti-human monoclonal antibody that recognizes the extracellular domain of PD-L1 (clone 5H1-A3)[1], followed by incubation with the anti-mouse secondary antibody conjugated with protein A-gold particles (5 nm). Each staining step was followed by five PBS washes and ten ddH_2O washes before contrast staining with 2% uranyl acetate.

The size and concentration of exosomes purified from cell culture supernatants or patients' plasma were determined using a NanoSight NS300 (Malvern Instruments), which is equipped with fast video capture and particle-tracking software.

For iodixanol density gradient centrifugation, exosomes harvested by differential centrifugation were loaded on top of a discontinuous iodixanol gradient (5%, 10%, 20% and 40%, made by diluting 60% OptiPrep aqueous iodixanol with 0.2 5M sucrose in 10 mM Tris) and centrifuged at 100,000*g* for 18 h at 4 °C (Beckman Coulter, Optima MAX-XP). Twelve fractions of equal volume were collected from the top of the gradients, with the exosomes distributed at the density range between 1.13 and 1.19 g/ml, as previously demonstrated[8,9,11,12]. The exosomes were further pelleted by ultracentrifugation at 100,000*g* for 2 h at 4 °C.

Immunoprecipitation. To analyse the role of ESCRT machinery in exosomal secretion of PD-L1 in melanoma cells, PD-L1/MEL624 cells were transfected with Flag-Hrs plasmid or vector and then lysed. The cleared lysate was incubated with Anti-FLAG Affinity Gel (Sigma-Aldrich) overnight at 4 °C. The immunoprecipitated proteins were resolved by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose membranes. PD-L1 and Flag (Hrs) were determined by western blot using specific antibodies.

ELISA. For detection of PD-L1 on extracellular vesicles, cell supernatants or patients' plasma, ELISA plates (96-well) (Biolegend) were coated with 0.25 μg per well (100 μl) of monoclonal antibody against PD-L1 (clone 5H1-A3) overnight at 4 °C. Free binding sites were blocked with 200 μl of blocking buffer (Pierce) for 1 h at room temperature. Then, 100 μl of plasma samples with or without extracellular vesicle removal, or extracellular vesicle samples purified from plasma or cell culture supernatants, were added to each well. The exosome or microvesicle samples purified from cell culture supernatants were prepared by serial dilution according to the total protein level to analyse the enrichment of PD-L1 on exosomes and microvesicles. The concentration of PD-L1 on the surface of exosomes isolated from indicated cells was calculated based on the linear range of the ELISA assay data. The exosome or microvesicle samples derived from the plasma samples of healthy donors or melanoma patients were prepared using the same volume of PBS as the plasma as they were originally derived from. The plasma samples with (extracellular vesicle-excluded) or without (total) extracellular vesicle removal were diluted with PBS in a 1:0.75 volume ratio. After overnight incubation at 4 °C, biotinylated monoclonal PD-L1 antibody (clone MIH1, eBioscience) was added to each well and incubated for 1 h at room temperature. A total of 100 μl per well of horseradish peroxidase-conjugated streptavidin (BD Biosciences) diluted in PBS containing 0.1% BSA was then added and incubated for 1 h at room temperature. Plates were developed with tetramethylbenzidine (Pierce) and stopped with 0.5N H_2SO_4. The plates were read at 450 nm with a BioTek plate reader. Recombinant human PD-L1 protein (R&D Systems, Cat# 156-B7) was used to make a standard curve. Recombinant P-selectin protein (R&D Systems, Cat# 137-PS) was used as negative control to verify the detection specificity. The result of standard curve demonstrated that the established ELISA exhibited a reliable linear detection range from 0.2 to 12 ng/ml.

For detection of IFN-γ, TNF and IL-2, the supernatant of human CD8 T cells was harvested and measured according to the kit manufacturer's instructions (Biolegend).

PD-1–PD-L1 binding assay. To test the binding of exosomal PD-L1 to PD-1, 100 μl of exosome samples of different concentrations were captured onto PD-L1 antibody (clone 5H1-A3)-coated 96-well ELISA plates by overnight incubation at 4 °C. Then 100 μl of 4 μg/ml biotin-labelled human PD-1 protein (BPS Bioscience, Cat# 71109) was added and incubated for 2 h at room temperature. A total of 100 μl per well of horseradish peroxidase-conjugated streptavidin (BD Biosciences) diluted in PBS containing 0.1% BSA was then added and incubated for 1 h at room temperature. Plates were developed with tetramethylbenzidine (Pierce) and stopped using 0.5N H_2SO_4. The plates were read at 450 nm with a BioTek plate reader. Recombinant human PD-L1 protein directly coated onto the plates was used as the positive control.

Treatment of CD8 T cells with the exosomes. To block PD-L1 on the exosome surface, purified exosomes (200 μg) were incubated with PD-L1 blocking antibodies (10 μg/ml) or IgG isotype antibodies (10 μg/ml) in 100 μl PBS, and then washed with 30 ml PBS and pelleted by ultracentrifugation to remove the non-bound free antibodies. Human CD8 T cells purified from peripheral blood using

immunodepletion on a Ficoll-Hypaque gradient (RosetteSep, StemSep Technologies) or mouse CD8 T cells purified from splenocytes using Dynabeads Untouched Mouse CD8 Cells Kit (Invitrogen) were stimulated with anti-CD3 (2 μg/ml) and anti-CD28 (2 μg/ml) antibodies for 24 h and then incubated with human melanoma cell-derived exosomes or mouse B16-F10 cell-derived exosomes with or without PD-L1 blocking for 48 h in the presence of anti-CD3/CD28 antibodies. For human CD8 T cells (2×10^5 cells/well in a 96-well plate), 25 μg/ml of human WM9 cell-derived exosomes (carrying surface PD-L1 at a level of 0.05 ng per μg of exosomes as determined by ELISA, Fig. 1i) were used as the circulating exosomal PD-L1 level in melanoma patients is around 1.25 ng/ml (Fig. 2h). For mice CD8 T cells (2×10^5 cells/well in a 96-well plate), 100 μg/ml of mouse B16-F10 cell-derived exosomes (carrying surface PD-L1 at a level of 0.016 ng per μg of exosomes as determined by ELISA) were used as the circulating exosomal PD-L1 level in mice bearing B16-F10 tumours is around 1.63 ng/ml. The treated cells were then collected, stained, and analysed by flow cytometry. Information about the primary antibodies is included in Supplementary Table 2. To assay for the proliferation of CD8 T cells, CFSE, a dye for the tracking of cell division (Molecular Probes) was used. A total of 1×10^6 CD8 T cells were stained with CFSE at 5 μM. The cells were then incubated at 37 °C for 20 min and the reaction was stopped by adding 5 volumes of cold medium with 10% FBS, and treated as above. Unstimulated CFSE-labelled cells served as a non-dividing control.

The exosome–T cell binding assay. To verify the physical interactions between melanoma cell-derived exosomes and CD8 T cells, purified exosomes were stained with CFSE in 100 μl PBS, and then washed with 10 ml PBS and pelleted by ultracentrifugation. Unstimulated or stimulated human CD8 T cells (2×10^5 cells/well in 96-well plates) were treated with CFSE-labelled exosomes (25 μg/ml) for 2 h, and then fixed for flow cytometry or confocal microscopy after immunostaining for CD8 T cells.

Generation of dendritic cells from bone marrow. Dendritic cells (DCs) were generated from bone marrow of C57BL/6 mice and cultured in RPMI 1640 with 10% (v/v) FBS, 20 mM L-glutamine, 50 μM β-mercapoethanol, 20 ng/ml IL-4 and 20 ng/ml GM-CSF. After 3 days, half of the culture medium was replaced by fresh medium containing 40 ng/ml IL-4 and 40 ng/ml GM-CSF. To prime antigen-specific OT-I CD8 T cells, DCs were subsequently loaded with 2 μg/ml SIINFEKL ($OVA_{257–264}$) peptide overnight.

CD8 T cell-mediated tumour cell killing assay. To determine the effects of melanoma cell-derived exosomes on the ability of CD8 T cells to kill tumour cells, CD8 T cells were purified from the splenocytes of OT-I mice expressing a transgene encoding a T cell receptor that specifically recognized SIINFEKL peptide bound to MHC-I H-$2k^{b}$[32]. OT-I CD8 T cells (4×10^5 cells/well in a 48-well plate) were then activated by incubation with SIINFEK-loaded (2 μg/ml) bone marrow-derived DCs (2×10^5 cells/well). The activated OT-I CD8 T cells (4×10^5 cells/well in 48-well plate) were treated with PBS (as a control) or B16-F10-derived exosomes (100 μg/ml for 48 h) with or without IgG isotype or PD-L1 antibody blocking (10 μg/ml), and then co-cultured with CFSE-labelled melanoma PD-L1 (KD) B16/OVA cells (4×10^5) in 6-well plates for 48 h at an effector to target (E:T) ratio of 1:1. Cells were then harvested, intracellularly stained with BV650-conjugated antibody against cleaved-caspase-3 (BD Biosciences) and analysed by flow cytometry. Information about the primary antibodies is included in Supplementary Table 2.

Immunofluorescence staining. Immunofluorescence staining was performed on fixed cells or formalin-fixed, paraffin-embedded (FFPE) sections. For fixed cells, permeabilization with 0.1% Triton X-100 was performed before blocking with bovine serum albumin (BSA) buffer for 1 h. For FFPE sections, antigen retrieval by steaming in citrate buffer (pH = 6.0) was performed before blocking. The fixed cells or FFPE sections were incubated with primary antibodies overnight at 4 °C, followed by incubation with fluorophore-conjugated secondary antibodies for 1 h. Nuclei were stained with DAPI. Samples were observed using a Nikon confocal microscope at 100× magnification.

Western blot analysis. Whole cell lysates or exosomal proteins were separated using 12% SDS–PAGE and transferred onto nitrocellulose membranes. The blots were blocked with 5% non-fat dry milk at room temperature for 1 h, and incubated overnight at 4 °C with the corresponding primary antibodies at dilutions recommended by the suppliers, followed by incubation with HRP-conjugated secondary antibodies (Cell Signaling Technology) at room temperature for 1 h. The blots on the membranes were developed with ECL detection reagents (Pierce). CD63, Hrs, Alix, and TSG101 were used as exosome markers. TYRP-1 and TYRP-2 were used as melanoma-specific markers. GAPDH was used as a loading control. Information about the primary antibodies was included in Supplementary Table 2.

Quantitative PCR (qPCR). Total RNA was isolated from CD8 T cells using TRIzol Reagent (Invitrogen), and reverse transcribed into first-strand complementary DNA (cDNA) with random primer with RevertAid First Strand cDNA Synthesis Kit (ThermoFisher Scientific). The samples were then analysed in an Applied Biosystems QuantStudio 3 Real-Time PCR system. *GAPDH* was used as an internal control. Information about the primers is included in Supplementary Table 3.

In vivo mice study. All animal experiments were performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Pennsylvania. For establishing human melanoma xenograft model in nude mice, WM9 cells (5×10^6 cells in 100 μl medium) were injected into flanks of 8-week-old female athymic nude mice. Tumours were measured using a digital caliper and the tumour volume was calculated by the formula: $(\text{width})^2 \times \text{length}/2$. Mice were euthanized 30 days after cell inoculation or if the longest dimension of the tumours reached 2.0 cm before 30 days. Immediately following euthanasia, blood samples were harvested by cardiac puncture, and exosomes were purified and detected by ELISA using the aforementioned method. Exosomes purified from sex-, age- and weight-matched healthy nude mice without xenograft were used as the control.

For establishing syngeneic mouse melanoma model in C57BL/6 mice, B16-F10 cells or B16-F10 PD-L1 (KD) cells (5×10^5 cells in 100 μl medium) were subcutaneously injected into immunocompetent C57BL/6 mice. Based on the difference in the level of circulating exosomal PD-L1 between mice bearing parental B16-F10 and PD-L1 (KD) B16-F10 tumours (1.63 ng/ml vs 0.70 ng/ml), a total of 100 μg of parental B16-F10 cell-derived exosomes (carrying surface PD-L1 at a level of 0.016 ng per μg of exosomes) with or without IgG isotype, CD63 or PD-L1 blocking (10 μg/ml) were injected into mice after inoculation of PD-L1 (KD) B16-F10 cells to examine the functional significance of PD-L1. The dose of 100 μg exosomes used for our in vivo study was equivalent to approximately 30% of the physiological level of circulating exosomes in mice, and was also comparable to those from a palpable tumour in mice according to our data. Tail vein injections of exosomes (100 μg in 100 μl PBS) were performed every 3 days. Mice were weighed every 3 days. Tumours were measured using a digital caliper and the tumour volume was calculated by the formula: $(\text{width})^2 \times \text{length}/2$. The mice were euthanized before the longest dimension of the tumours reached 2.0 cm. Mice were allocated randomly to each treatment group. Downstream analyses of mouse samples (immunofluorescence staining, flow cytometry and ELISA) were performed in a blinded fashion. For flow cytometry, the spleen and tumour samples were harvested, and single cell suspensions were prepared and red blood cells were lysed using ACK Lysis Buffer. Information about the primary antibodies is included in Supplementary Table 2.

Reverse phase protein array (RPPA). RPPA was performed at the MD Anderson Cancer Center core facility using 50 μg protein per sample. All of the antibodies were validated by western blot[13]. Methods for data analysis are described below.

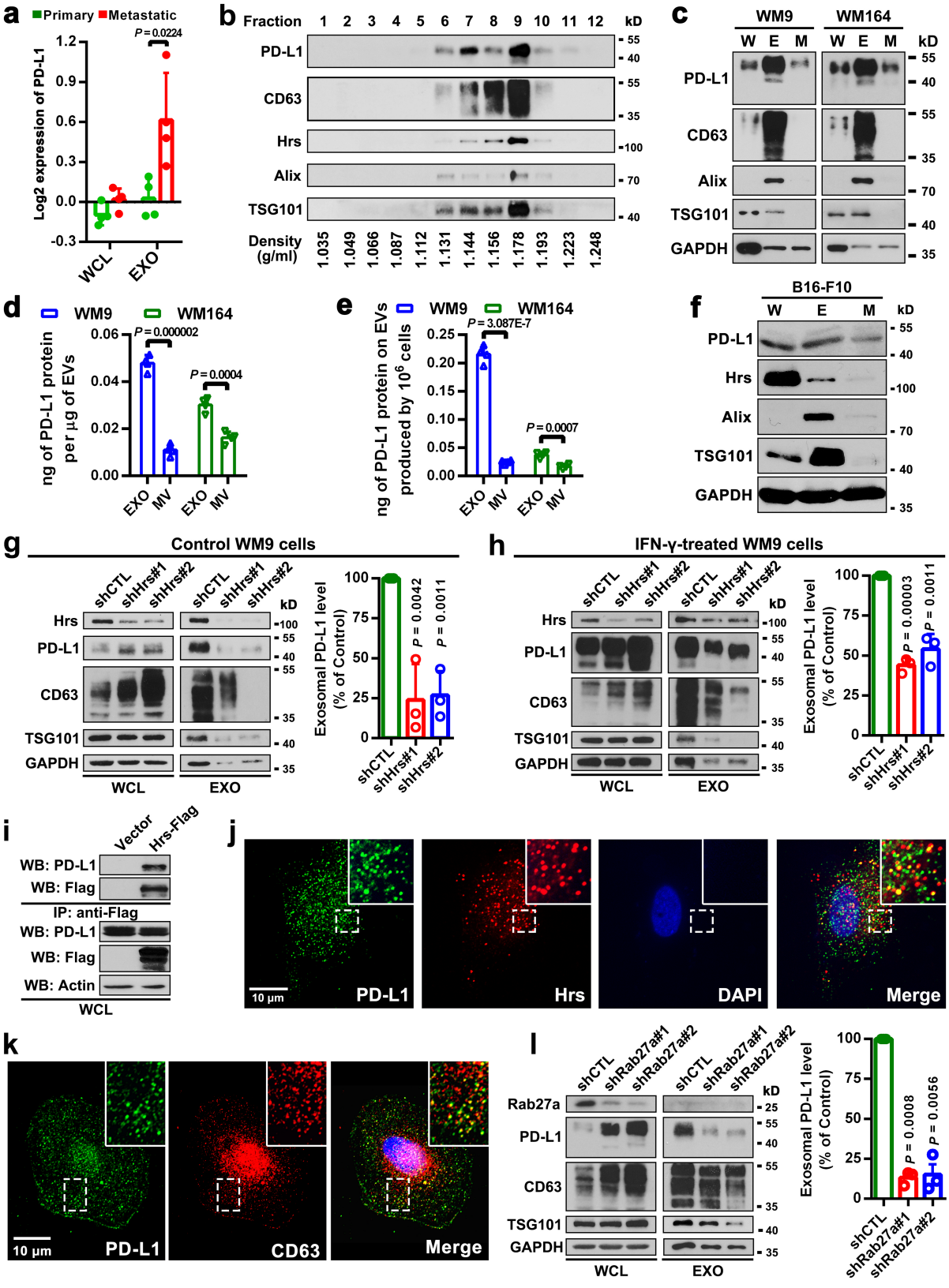
Statistical analyses. RPPA data analysis was performed according to the protocol from the MD Anderson Cancer Center. Specifically, relative protein levels for each sample were determined by interpolation of each dilution curve from the 'standard curve' (supercurve) of the slide (antibody). Supercurve is constructed by a script in R written by the RPPA core facility. The package binaries of SuperCurve and SuperCurveGUI are available in R-Forge (https://r-forge.r-project.org/R/?group_id=1899). These values are defined as supercurve $\log_2$ value. All the data points were normalized for protein loading and transformed to linear value, designated as 'normalized linear'. Normalized linear value was transformed to the $\log_2$ value, and then median-centred for further analysis. Median-centred values were obtained by subtracting the median of all samples in a given protein. All of the above-mentioned procedures were performed by the RPPA core facility. The normalized data provided by the RPPA core facility were analysed by Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/) and visualized using the Java TreeView 1.0.5 (http://jtreeview.sourceforge.net/).

All other statistical analyses were performed using GraphPad Prism v.6.0. Normality of distribution was determined by D'Agostino–Pearson omnibus normality test and variance between groups was assessed by the *F*-test. For normally distributed data, significance of mean differences was determined using two-tailed paired or unpaired Student's *t*-tests; for groups that differed in variance, unpaired *t*-test with Welch's correction was performed. For data that were not normally distributed, non-parametric Mann–Whitney *U*-tests or Wilcoxon matched-pairs tests were used for unpaired and paired analysis, respectively. Correlations were determined by Pearson's *r* coefficient. Two-way ANOVA was used to compare mouse tumour volume data among different groups. log-rank and Wilcoxon tests were used to analyse the mouse survival data. Error bars shown in graphical data represent mean ± s.d. A two-tailed value of $P < 0.05$ was considered statistically significant.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

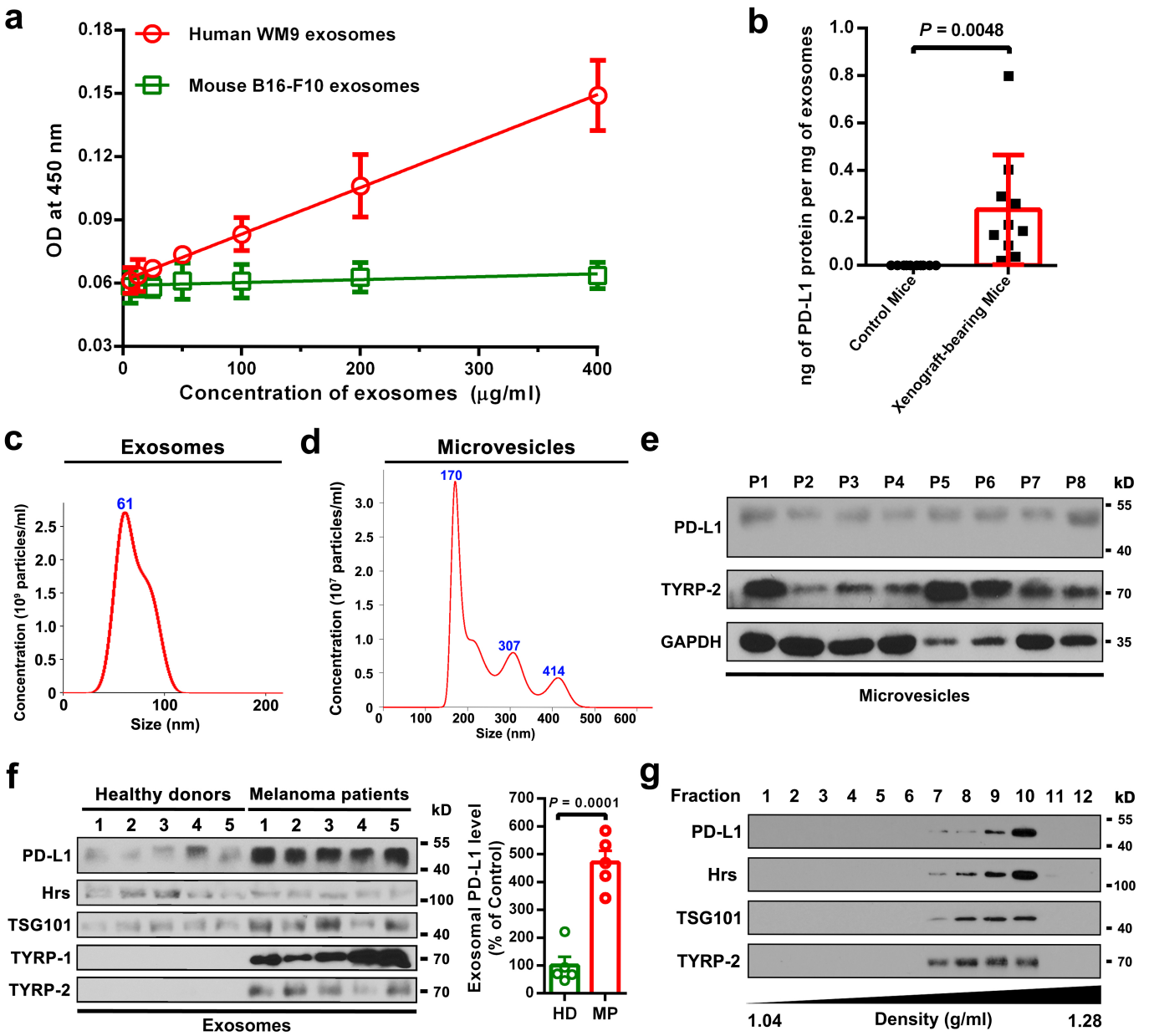
Data availability. All data and materials are available from the authors upon reasonable request.

31. Nishino, M. et al. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. *Clin. Cancer Res.* **19**, 3936–3943 (2013).
32. Hogquist, K. A. et al. T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).

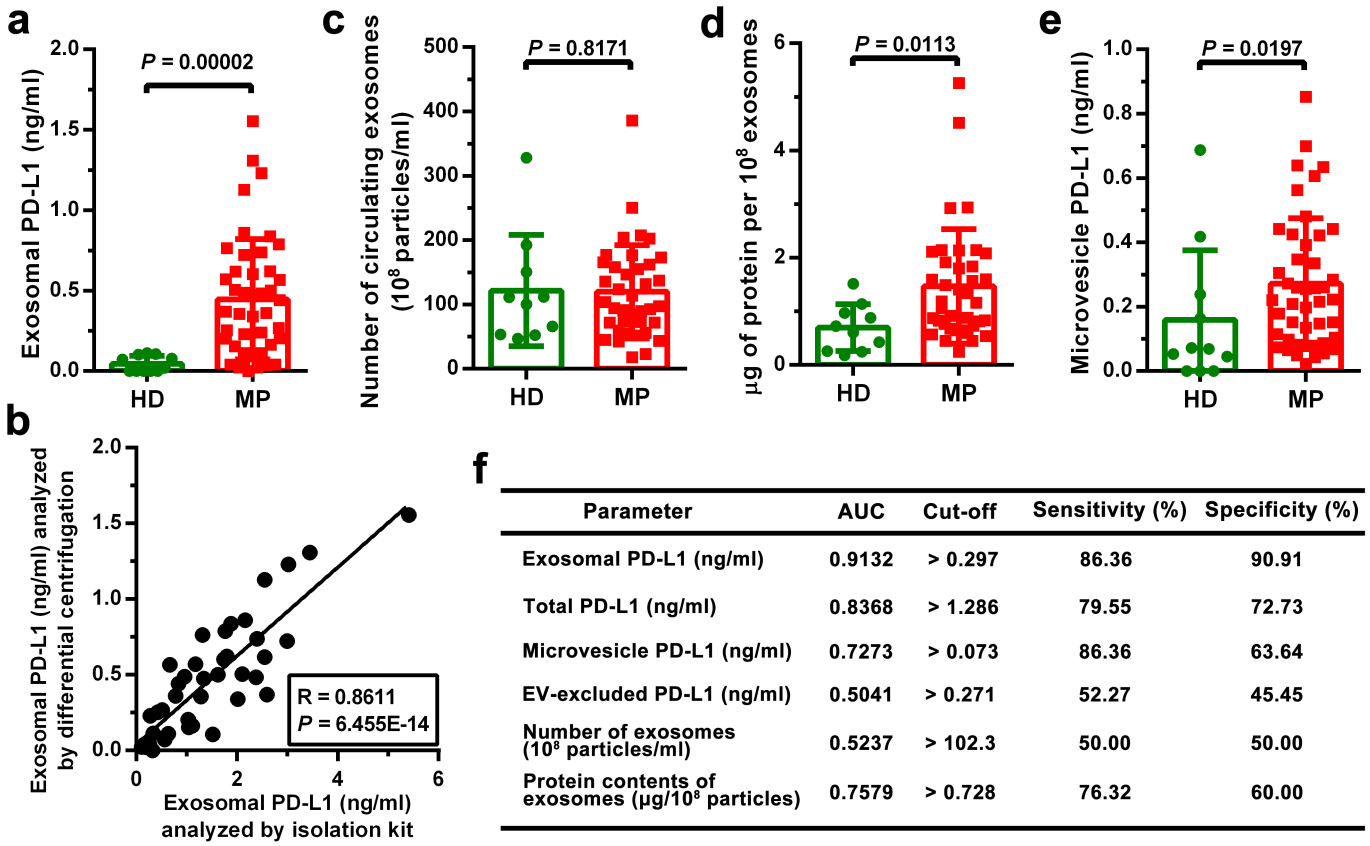


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Melanoma cells release extracellular vesicles carrying PD-L1. a, The $\log_2$-transformed RPPA data showing a higher level of exosomal PD-L1 secreted by metastatic melanoma cell lines compared with primary melanoma cell lines. Data represent mean ± s.d. of four primary (WM1552C, WM35, WM793, WM902B) or metastatic (UACC-903, 1205Lu, WM9, WM164) melanoma lines. **b**, Density gradient centrifugation confirming that PD-L1 secreted by WM9 cells co-fractionated with exosome markers CD63, Hrs, Alix and TSG101. **c**, Immunoblots for PD-L1 in the whole cell lysate (W), purified exosomes (E) or microvesicles (M) from different metastatic melanoma cell lines. The same amount of protein was loaded in each lane. **d**, Levels of PD-L1 on the exosomes or microvesicles derived from melanoma cells as assayed by ELISA. **e**, The levels of exosomal PD-L1 and microvesicle PD-L1 produced by an equal number of melanoma cells. **f**, Immunoblots for PD-L1 in the whole cell lysate, purified exosomes or microvesicles from mouse melanoma B16-F10 cells. The same amount of protein was loaded in each lane. **g**, **h**, Western blot analysis of PD-L1 in *Hrs* knockdown cells without (**g**) or with (**h**) IFN-γ treatment. Quantification of the western blotting data (**g**, right; **h**, right). **i**, Co-immunoprecipitation of PD-L1 and Hrs from MEL624 cells expressing exogenous PD-L1 and Hrs. **j**, Immunofluorescence staining of intracellular PD-L1 and exosome marker Hrs in WM9 cells treated with IFN-γ. **k**, Immunofluorescence staining of intracellular PD-L1 and CD63 in WM9 cells treated with IFN-γ. **l**, western blotting analysis showing intracellular accumulation of PD-L1, and decreased exosomal secretion of PD-L1 in WM9 cells with *RAB27A* knockdown (left). The levels of exosomal PD-L1 were compared (right). Two experiments were repeated independently with similar results (**b**, **c**, **f**, **i**–**k**). Data represent mean ± s.d. of four (**d**, **e**), or three (**g**, **h**, **l**) independent biological replicates. Statistical analysis is performed by two-sided unpaired *t*-test (**a**, **d**, **e**, **g**, **h**, **l**). For gel source data (**b**, **c**, **f**–**i**, **l**), see Supplementary Fig. 1.



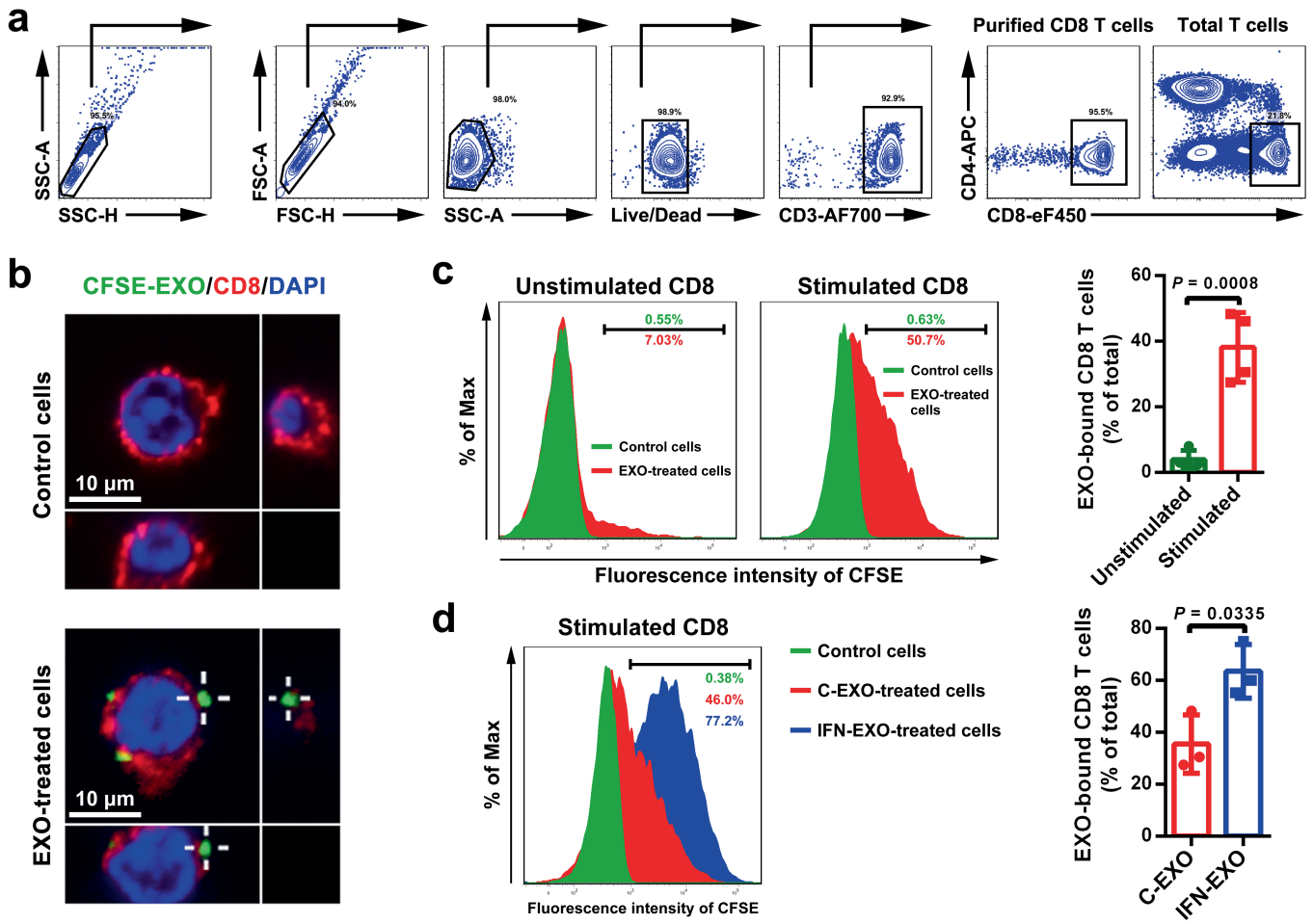
Extended Data Fig. 2 | Melanoma cells secrete exosomal PD-L1 into the circulation. a, The monoclonal antibodies against the extracellular domain of human PD-L1 specifically detect human exosomal PD-L1, but not mouse exosomal PD-L1 ($n = 3$ biologically independent experiments). **b**, Levels of human PD-L1 in exosomes from the plasma of control nude mice ($n = 10$) and human WM9 melanoma xenograft-bearing nude mice ($n = 10$) per mg of total circulating exosomal proteins. **c**, Characterization of circulating exosomes purified from the plasma of a patient with Stage IV melanoma using NanoSight nanoparticle tracking analysis. **d**, Characterization of circulating microvesicles purified from the plasma sample of a patient with Stage IV melanoma using NanoSight nanoparticle tracking analysis. **e**, Immunoblots for PD-L1 in the microvesicles purified from the plasma samples of 8 patients with Stage IV melanoma (denoted as P1–P8). **f**, Immunoblots for PD-L1 in the exosomes purified from the plasma samples of 5 healthy donors and 5 patients with stage IV melanoma (left panel). Quantification of the levels of exosomal PD-L1 by western blot analysis (right panel). Results are expressed as the percentage of the mean value of healthy donors. **g**, Standard density gradient centrifugation analysis showing that circulating PD-L1 co-fractionated with exosome markers Hrs and TSG101 and melanoma-specific marker TYRP-2. Three (**c**, **d**) or two (**e**, **g**) experiments were repeated independently with similar results. Data represent mean ± s.d. (**a**, **b**, **f**). Statistical analyses were performed using two-sided unpaired *t*-test (**b**, **f**). For gel source data (**e**–**g**), see Supplementary Fig. 1.



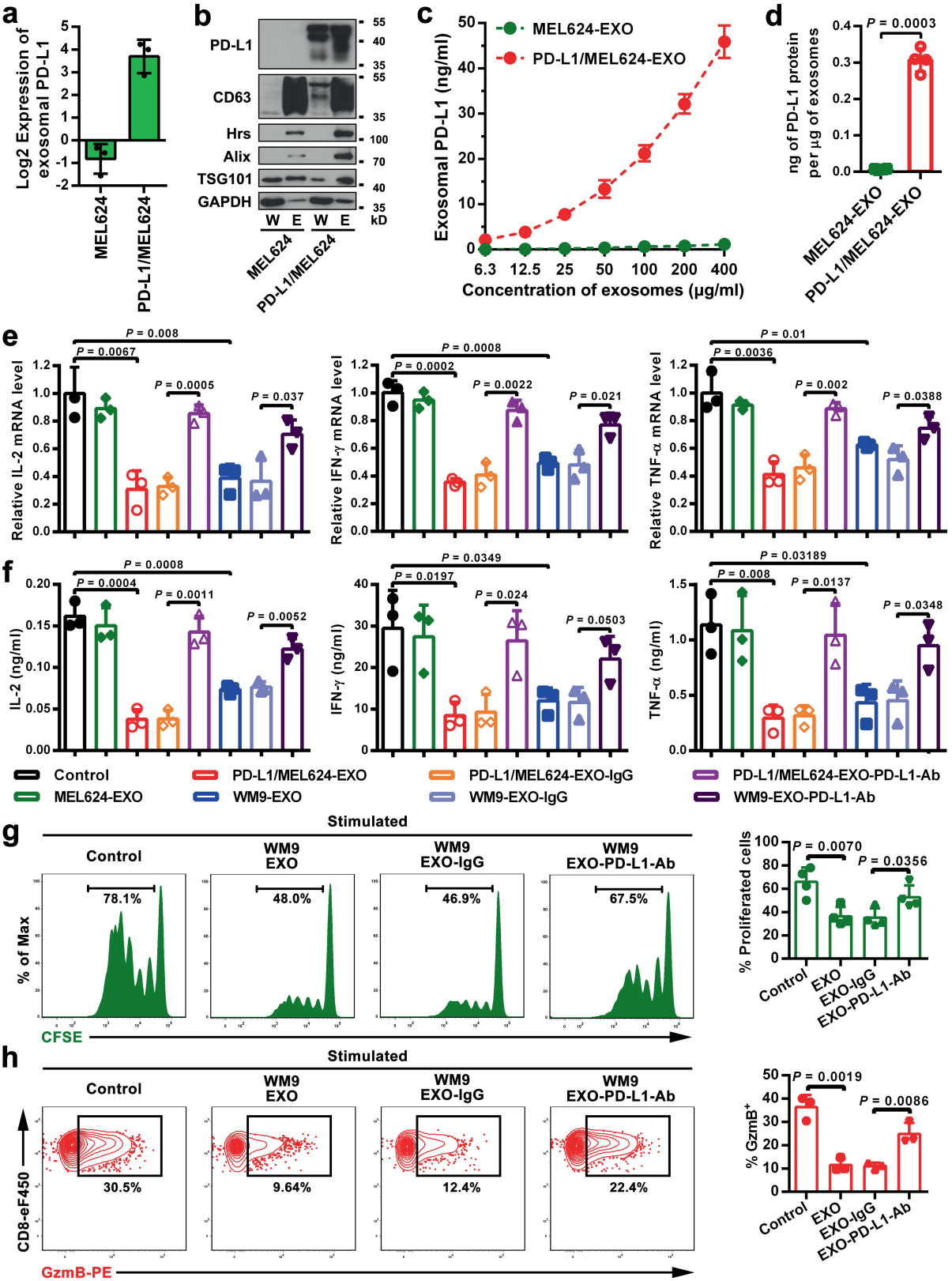
f

Parameter	AUC	Cut-off	Sensitivity (%)	Specificity (%)
Exosomal PD-L1 (ng/ml)	0.9132	> 0.297	86.36	90.91
Total PD-L1 (ng/ml)	0.8368	> 1.286	79.55	72.73
Microvesicle PD-L1 (ng/ml)	0.7273	> 0.073	86.36	63.64
EV-excluded PD-L1 (ng/ml)	0.5041	> 0.271	52.27	45.45
Number of exosomes (10^8 particles/ml)	0.5237	> 102.3	50.00	50.00
Protein contents of exosomes (μg/10^8 particles)	0.7579	> 0.728	76.32	60.00

Extended Data Fig. 3 | The number or bulk protein level of circulating exosomes shows no or modest difference between healthy donors and patients with metastatic melanoma. a, ELISA showing the level of PD-L1 on circulating exosomes purified from healthy donors (HD, $n = 11$) and melanoma patients (MP, $n = 44$). The exosomes were purified using differential centrifugation. **b**, Pearson correlation between the ELISA-detected levels of PD-L1 on circulating exosomes purified by differential centrifugation or using the commercial exosome isolation kit ($n = 44$). **c**, Comparison of the number of circulating exosomes between healthy donors ($n = 10$) and melanoma patients ($n = 38$). **d**, Comparison of the protein content of circulating exosomes between healthy donors ($n = 10$) and melanoma patients ($n = 38$). **e**, ELISA of the circulating level of microvesicle PD-L1 in healthy donors (HD, $n = 11$) and melanoma patients (MP, $n = 44$). **f**, Detailed data associated with the ROC curve analysis depicted in Fig. 2g. Data represent mean ± s.d. Statistical analyses are performed by two-sided unpaired *t*-test (**a**, **c**-**e**).

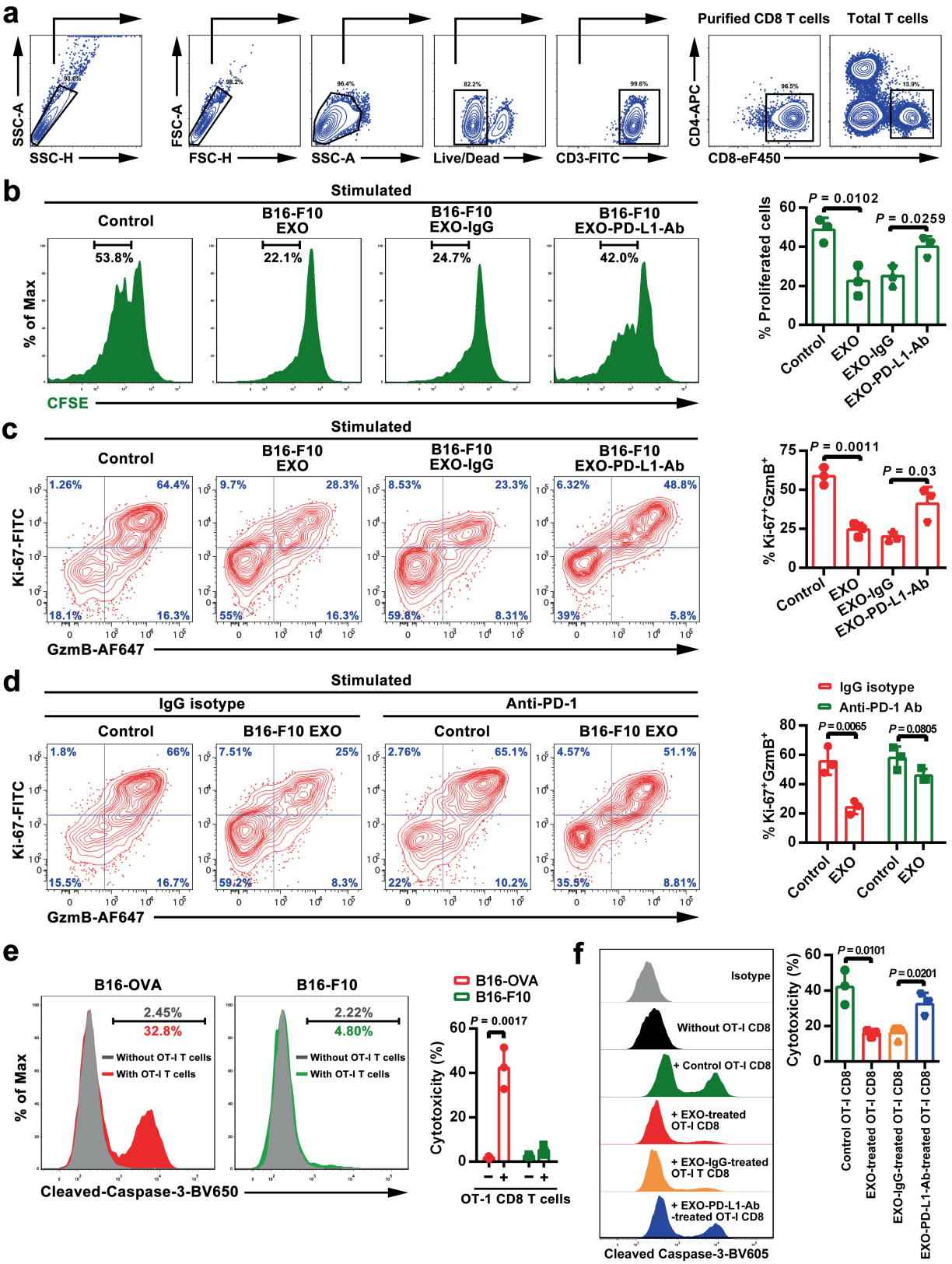


Extended Data Fig. 4 | Melanoma cell-derived exosomes bind to CD8 T cells on their surface. a, Representative contour plots showing the general gating strategy used to identify the purified CD8 T cells ($CD3^+CD8^+CD4^-$) from human peripheral blood. **b**, Confocal microscopy analysis of human peripheral CD8 T cells (stimulated with anti-CD3/CD28 antibodies) after incubation with CFSE-labelled WM9 cell-derived exosomes for 2 h. The experiments were repeated three times independently with similar results. **c**, Representative histogram of human peripheral CD8 T cells with or without anti-CD3/CD28 antibody stimulation after incubation with CFSE-labelled WM9 cell-derived exosomes for 2 h (left). The proportion of exosome-bound cells is shown (right). **d**, Representative histogram of human peripheral CD8 T cells (stimulated with anti-CD3/CD28 antibodies) after incubation with the same number of CFSE-labelled exosomes purified from control or IFN-γ-treated WM9 cells for 2 h (left panel). The proportion of EXO-bound cells is shown in the right panel. Data represent mean ± s.d. of four (**c**) or three (**d**) independent biological replicates. Statistical analyses are performed using two-sided unpaired *t*-test (**c**, **d**).



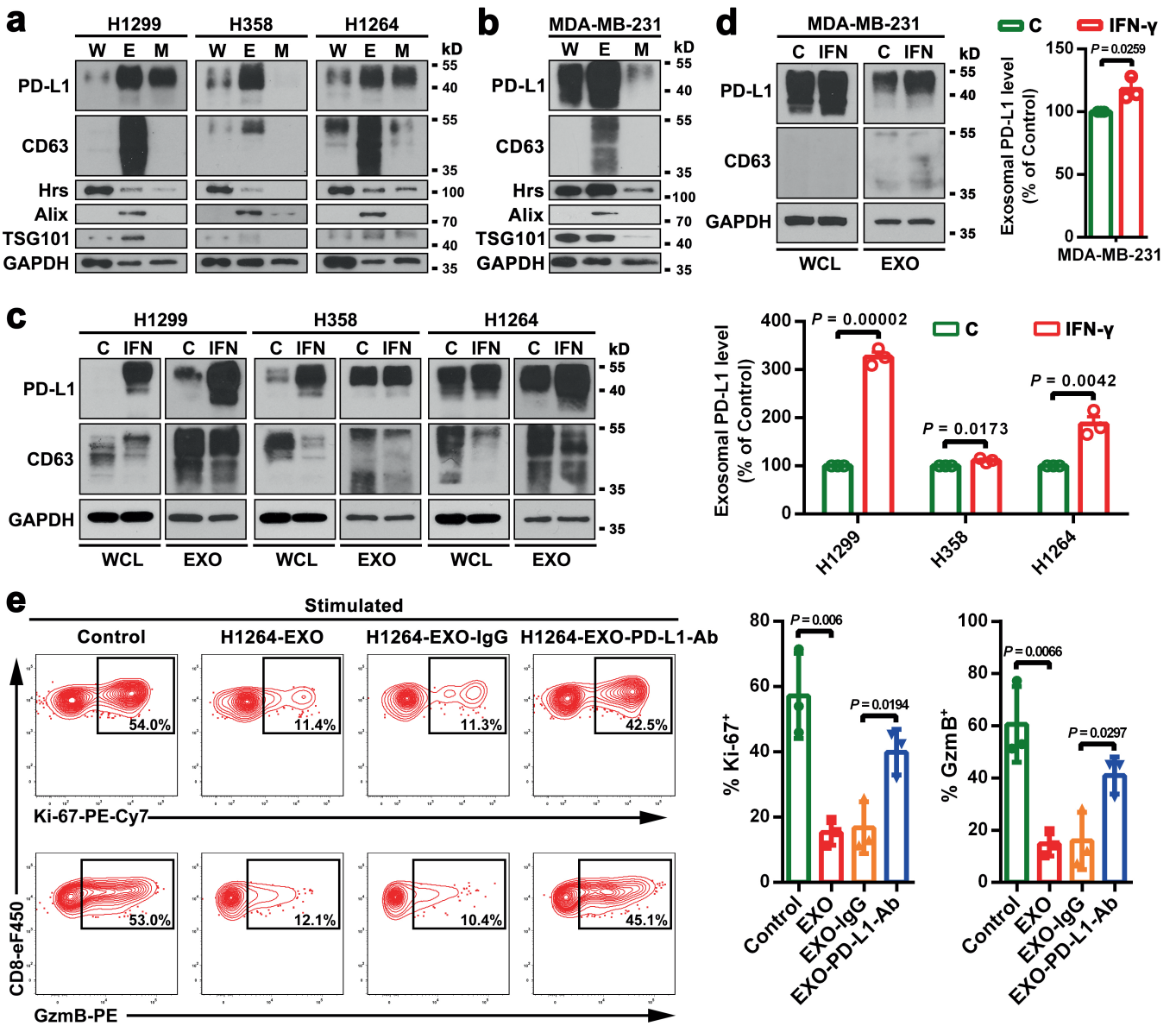
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Functional inhibition of CD8 T cells by exosomal PD-L1. a, The log_2-transformed RPPA data showing the levels of PD-L1 in the exosomes secreted by control (MEL624) or PD-L1-expressing (PD-L1/MEL624) human melanoma MEL624 cells (Bottom). **b**, Immunoblots for PD-L1 in the whole cell lysate (W) or in the purified exosomes (E) from MEL624 or PD-L1/MEL624 cells. The same amount of protein was loaded in each lane. The experiments were repeated two times independently with similar results. For source data, see Supplementary Fig. 1. **c**, PD-L1 on the surface of exosomes secreted by MEL624 or PD-L1/MEL624 cells as determined by ELISA. **d**, Levels of PD-L1 on exosomes secreted by MEL624 or PD-L1/MEL624 cells, as measured by ELISA. **e**, qPCR analyses of IL-2, IFN-γ, and TNF in human peripheral CD8 T cells (stimulated with anti-CD3/CD28 antibodies) after treatment with MEL624 cell-derived exosomes, PD-L1/MEL624 cell-derived exosomes or WM9-cell-derived exosomes with or without blocking by IgG isotype or the anti-PD-L1 antibodies. The relative mRNA expression level was calculated as the ratio to the control cells. **f**, ELISA of IL-2, IFN-γ, and TNF in human peripheral CD8 T cells (stimulated with anti-CD3/CD28 antibodies) after treatment with MEL624 cell-derived exosomes, PD-L1/MEL624 cell-derived exosomes or WM9-cell-derived exosomes with or without blocking by IgG isotype or PD-L1 antibodies. **g**, Representative histogram of CFSE-labelled human peripheral CD8 T cells (stimulated with anti-CD3/CD28 antibodies) after treatment with WM9 cell-derived exosomes with or without antibody blocking (left). The proportion of cells with diluted CFSE dye is shown (right). **h**, Representative contour plots of human peripheral CD8 T cells (stimulated with anti-CD3/CD28 antibodies) examined for the expression of granzyme B (GzmB) after treatment with WM9 cell-derived exosomes with or without antibody blocking (left). The percentage of $GzmB^+$ CD8 T cells stimulated with anti-CD3/CD28 antibodies is shown at the right panel. Data represent mean ± s.d. of three (**a**, **c**, **e**, **f**, **h**) or four (**d**, **g**) independent biological replicates. Statistical analyses are performed using two-sided unpaired *t*-test (**d**–**h**).

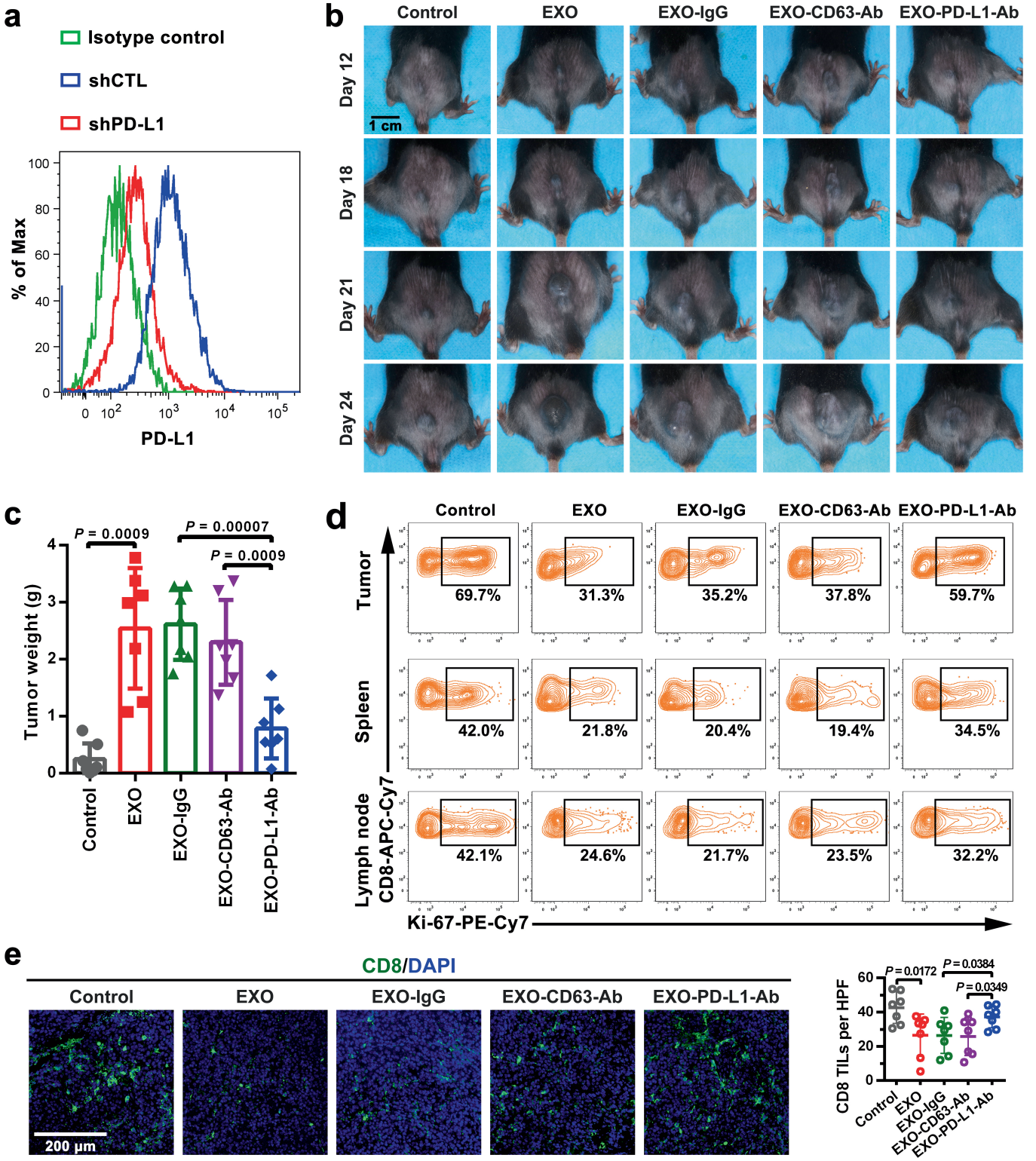


Extended Data Fig. 6 | See next page for caption.

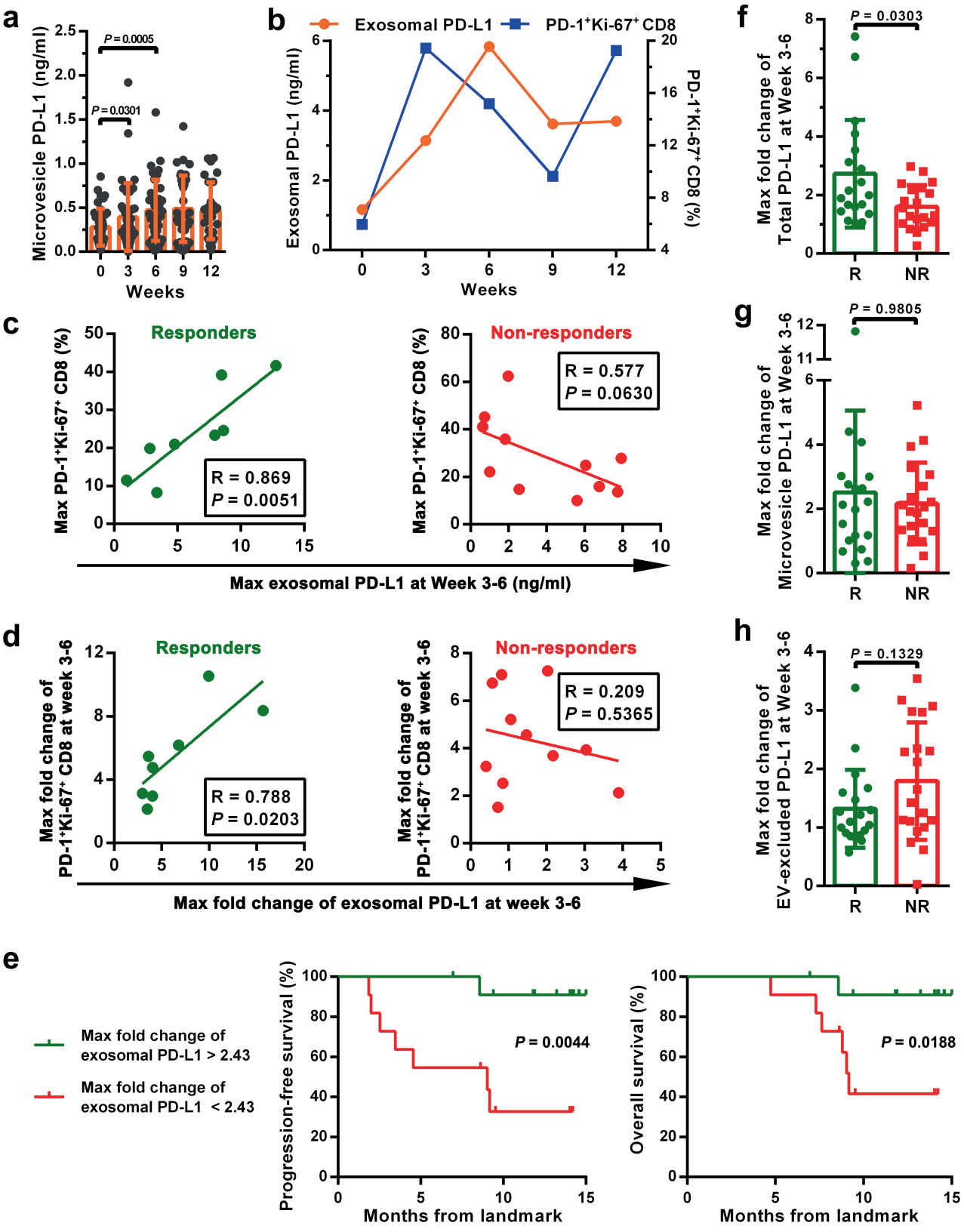
Extended Data Fig. 6 | Exosomal PD-L1 secreted by mouse melanoma B16-F10 cells inhibits the proliferation and cytotoxicity of mouse splenic CD8 T cells. a, Representative contour plots showing the general gating strategy used to identify the purified CD8 T cells ($CD3^+CD8^+CD4^-$) from mouse splenocytes. **b**, Representative histogram of CFSE-labelled mouse splenic CD8 T cells (stimulated with anti-CD3/CD28 antibodies) after treatment with B16-F10 cell-derived exosomes with or without blocking by IgG isotype or the anti-PD-L1 antibodies (left). The proportion of cells with diluted CFSE dye is shown at the right panel. **c**, Representative contour plots of mouse splenic CD8 T cells (stimulated with anti-CD3/CD28 antibodies) examined for the expression of Ki-67 and granzyme B (GzmB) after treatment with B16-F10 cell-derived exosomes with or without blocking by IgG isotype or the anti-PD-L1 antibodies (left). The percentage of $Ki\text{-}67^+GzmB^+$ CD8 T cells stimulated with anti-CD3/CD28 antibodies is shown (right). **d**, Representative contour plots of mouse splenic CD8 T cells (stimulated with anti-CD3/CD28 antibodies) examined for the expression of Ki-67 and GzmB after treatment with B16-F10 cell-derived exosomes in the presence or absence of anti-PD-1 blocking antibodies (left). The percentage of $Ki\text{-}67^+GzmB^+$ CD8 T cells stimulated with anti-CD3/CD28 antibodies is shown at the right panel. **e**, OT-I CD8 T cell-meditated tumour cell killing assay was performed in B16-OVA cells with *PD-L1* knockdown, or B16-F10 cells with *PD-L1* knockdown (negative control). Apoptosis of tumour cells was evaluated by flow cytometric analysis of intracellular cleaved caspase-3 (left), and the relative cytotoxicity was calculated (right). **f**, OT-I CD8 T cells, activated by OVA-pulsed bone marrow-derived dendritic cells and treated with PBS (as a control), exosomes derived from B16-F10 cells with or without IgG isotype or PD-L1 antibody blocking, were co-cultured with *PD-L1* knockdown B16-OVA cells for 48 h. Tumour cell apoptosis was evaluated by flow cytometric analysis of intracellular cleaved caspase-3 (left), and the relative cytotoxicity was calculated (right). Data represent mean ± s.d. of three (**b**–**f**) independent biological replicates. Statistical analyses are performed using two-sided unpaired *t*-test (**b**–**f**).



Extended Data Fig. 7 | Lung cancer and breast cancer cells release extracellular vesicles carrying PD-L1. a, Immunoblots for PD-L1 in the whole cell lysate (W), purified exosomes (E) or microvesicles (M) from different lung cancer cell lines. The same amounts of proteins were loaded for each fraction. **b**, Immunoblots for PD-L1 in the whole cell lysate, purified exosomes or microvesicles from the breast cancer cell line MDA-MB-231. The same amount of protein was loaded for each fraction. **c**, Immunoblots for PD-L1 in the whole cell lysate (WCL) or in the purified exosomes (EXO) from control (C) or IFN-γ-treated (IFN) lung cancer cells. The same amounts of exosome proteins from IFN-γ-treated and control cells were loaded (left). Quantification of the exosomal PD-L1 level determined by western blot analysis (right). **d**, Immunoblots for PD-L1 in the whole cell lysate or in the purified exosomes from control or IFN-γ-treated the breast cancer MDA-MB-231 cells. The same amounts of exosome proteins from IFN-γ-treated and control cells were loaded (left). Quantification of the exosomal PD-L1 level determined by western blot analysis (right). **e**, Representative contour plots of human peripheral CD8 T cells examined for the expression of Ki-67 and GzmB after treatment with H1264 cell-derived exosomes with or without blocking by IgG isotype or PD-L1 antibodies (left). The percentage of Ki-67$^+$ or GzmB$^+$ CD8 T cells is shown (right). The experiments were repeated twice independently with similar results (**a**, **b**). Data represent mean ± s.d. of three (**c**–**e**) independent biological replicates. Statistical analyses are performed using two-sided unpaired *t*-test (**c**–**e**). For source data (**a**–**d**), see Supplementary Fig. 1.

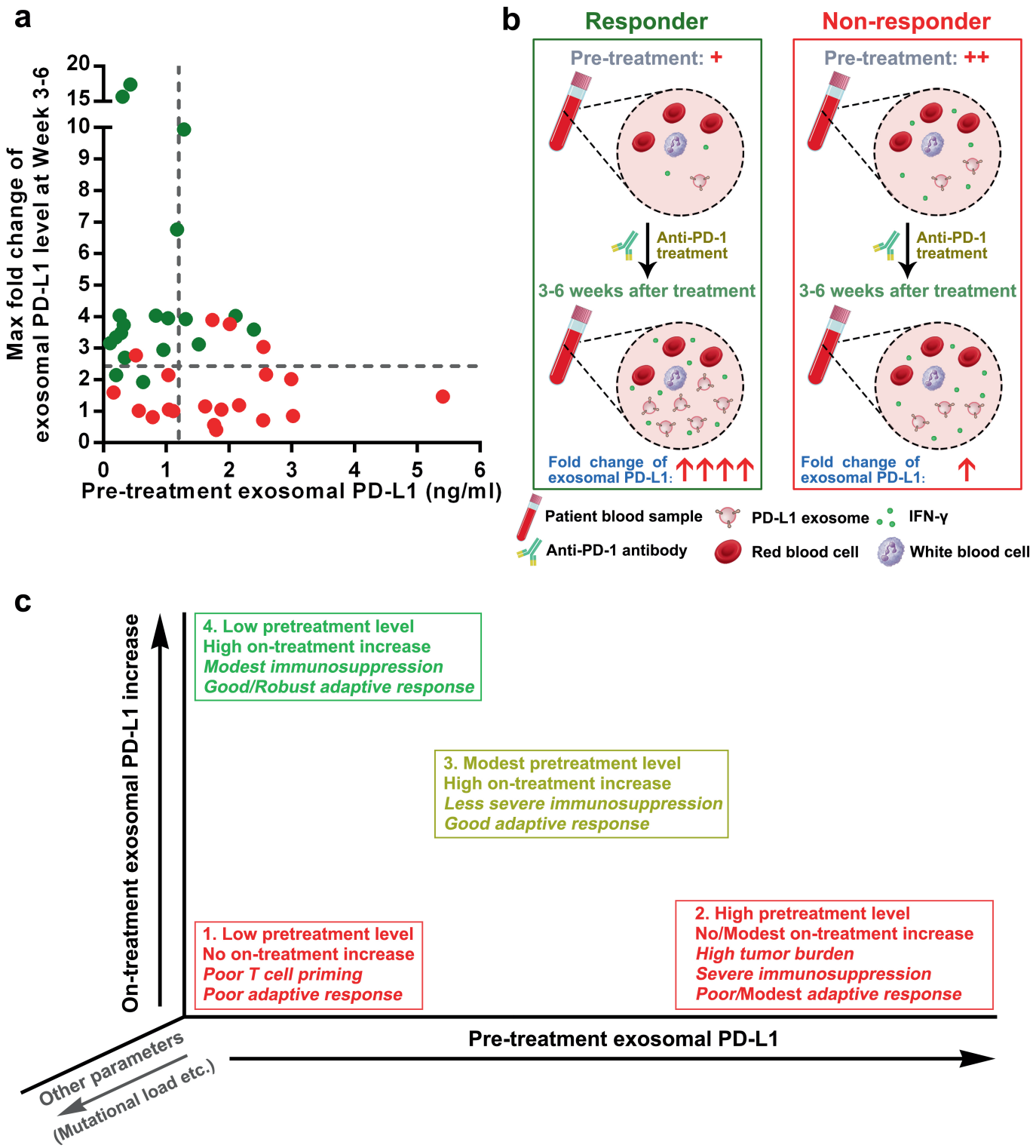


Extended Data Fig. 8 | Exosomal PD-L1 facilitates melanoma growth in vivo. a, Representative flow cytometric histograms of B16-F10 cells examined for the expression of PD-L1 with or without *PD-L1* knockdown. B16-F10 cells were stably depleted of *PD-L1* using lentiviral shRNA against *PD-L1* (shPD-L1) or the scrambled control shRNA (shCTL). The experiment was repeated twice independently with similar results. **b**, Representative images showing the growth of *PD-L1* knockdown B16-F10 tumours in C57BL/6 mice after indicated treatments. Experiments were performed using 7 mice for each group. **c**, The weights of *PD-L1* knockdown B16-F10 tumours from C57BL/6 mice with indicated treatments ($n = 7$ mice per group). Data represent mean ± s.d. **d**, Representative contour plot of CD8 TILs or splenic or lymph node CD8 T cells examined for the expression of Ki-67 after indicated treatments. Experiments were performed using 7 mice for each group. See Fig. 3c for quantification data. **e**, Representative immunofluorescence images of CD8 TILs in tumour tissues (left). The number of CD8 TILs for each mouse ($n = 7$ mice per group) were quantified from 5 high-power fields (HPF) (right). Statistical analysis is performed using two-sided unpaired *t*-test (**c**, **e**).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | The level of circulating exosomal PD-L1 distinguishes clinical responders to pembrolizumab treatment from non-responders. a, The levels of PD-L1 on circulating microvesicles at serial time points pre- and on-treatment ($n = 39$). **b**, The frequency of PD-1$^+$ Ki-67$^+$ CD8 T cells and the level of circulating exosomal PD-L1 in clinical responders at serial time points pre- and on-treatment ($n = 8$). **c**, Pearson correlation of the maximum level of circulating exosomal PD-L1 at week 3–6 to the maximum frequency of PD-1$^+$Ki-67$^+$ CD8 T cells at week 3–6 in clinical responders ($n = 8$) and non-responders ($n = 11$). **d**, Pearson correlation of the maximum fold change of circulating exosomal PD-L1 level at week 3–6 to the maximum fold change of PD-1$^+$Ki-67$^+$ CD8 T cells at week 3–6 in clinical responders ($n = 8$) and non-responders ($n = 11$). **e**, Kaplan–Meier progression-free and overall survival of patients with high ($n = 11$) and low ($n = 12$) fold changes of circulating exosomal PD-L1 at 3–6 weeks. **f**, Comparison of the maximum fold change of total circulating PD-L1 at week 3–6 between the clinical responders and non-responders. R, responders, $n = 19$; NR, non-responders, $n = 20$. **g**, Comparison of the maximum fold change of circulating microvesicle PD-L1 at week 3–6 between the clinical responders ($n = 19$) and non-responders ($n = 20$). **h**, Comparison of the maximum fold change of extracellular-excluded PD-L1 at week 3–6 between the clinical responders ($n = 19$) and non-responders ($n = 20$). Data represent mean ± s.d. Statistical analyses were performed using two-sided paired *t*-test (**a**), log-rank test (**e**), or two-sided unpaired *t*-test (**f**–**h**).



Extended Data Fig. 10 | Circulating exosomal PD-L1 is a potential rationale-based and clinically accessible predictor for clinical outcomes of anti-PD-1 therapy. a, Tracking the levels of circulating exosomal PD-L1 before and during anti-PD-1 treatment may stratify responders to anti-PD-1 therapy (green) from non-responders (red) as early as 3–6 weeks into the treatment. **b**, Potential application of circulating exosomal PD-L1 to predict patients' response to anti-PD-1 therapy. The pre-treatment level of circulating exosomal PD-L1 is lower in metastatic melanoma patients with clinical response to anti-PD-1 therapy. After 3–6 weeks of anti-PD-1 treatment, the level of circulating exosomal PD-L1 increases significantly in clinical responders but not in non-responders. **c**, Tracking both the pre-treatment and on-treatment levels of circulating exosomal PD-L1 may help identify the possible reasons for success (green) or failure (red) of the therapy.

LETTER

https://doi.org/10.1038/s41586-018-0290-0

A multiprotein supercomplex controlling oncogenic signalling in lymphoma

James D. Phelan[1,19], Ryan M. Young[1,19], Daniel E. Webster[1], Sandrine Roulland[1,18], George W. Wright[2], Monica Kasbekar[1], Arthur L. Shaffer III[1], Michele Ceribelli[3], James Q. Wang[1], Roland Schmitz[1], Masao Nakagawa[1], Emmanuel Bachy[1], Da Wei Huang[1], Yanlong Ji[4], Lu Chen[3], Yandan Yang[1], Hong Zhao[1], Xin Yu[1], Weihong Xu[1], Maryknoll M. Palisoc[5], Racquel R. Valadez[5], Theresa Davies-Hill[5], Wyndham H. Wilson[1], Wing C. Chan[6], Elaine S. Jaffe[5], Randy D. Gascoyne[7], Elias Campo[8], Andreas Rosenwald[9], German Ott[10], Jan Delabie[11], Lisa M. Rimsza[12], Fausto J. Rodriguez[13], Fayez Estephan[14], Matthias Holdhoff[14], Michael J. Kruhlak[15], Stephen M. Hewitt[16], Craig J. Thomas[1,3], Stefania Pittaluga[5], Thomas Oellerich[1,4,17,20]* & Louis M. Staudt[1,20]*

B cell receptor (BCR) signalling has emerged as a therapeutic target in B cell lymphomas, but inhibiting this pathway in diffuse large B cell lymphoma (DLBCL) has benefited only a subset of patients[1]. Gene expression profiling identified two major subtypes of DLBCL, known as germinal centre B cell-like and activated B cell-like (ABC)[2,3], that show poor outcomes after immunochemotherapy in ABC. Autoantigens drive BCR-dependent activation of NF-κB in ABC DLBCL through a kinase signalling cascade of SYK, BTK and PKCβ to promote the assembly of the CARD11–BCL10–MALT1 adaptor complex, which recruits and activates IκB kinase[4–6]. Genome sequencing revealed gain-of-function mutations that target the CD79A and CD79B BCR subunits and the Toll-like receptor signalling adaptor MYD88[5,7], with MYD88(L265P) being the most prevalent isoform. In a clinical trial, the BTK inhibitor ibrutinib produced responses in 37% of cases of ABC[1]. The most striking response rate (80%) was observed in tumours with both CD79B and MYD88(L265P) mutations, but how these mutations cooperate to promote dependence on BCR signalling remains unclear. Here we used genome-wide CRISPR–Cas9 screening and functional proteomics to determine the molecular basis of exceptional clinical responses to ibrutinib. We discovered a new mode of oncogenic BCR signalling in ibrutinib-responsive cell lines and biopsies, coordinated by a multiprotein supercomplex formed by MYD88, TLR9 and the BCR (hereafter termed the My-T-BCR supercomplex). The My-T-BCR supercomplex co-localizes with mTOR on endolysosomes, where it drives pro-survival NF-κB and mTOR signalling. Inhibitors of BCR and mTOR signalling cooperatively decreased the formation and function of the My-T-BCR supercomplex, providing mechanistic insight into their synergistic toxicity for My-T-BCR$^+$ DLBCL cells. My-T-BCR supercomplexes characterized ibrutinib-responsive malignancies and distinguished ibrutinib responders from non-responders. Our data provide a framework for the rational design of oncogenic signalling inhibitors in molecularly defined subsets of DLBCL.

We used a library of small guide RNAs (sgRNAs) to conduct genome-wide loss-of-function CRISPR–Cas9 screens for essential genes in lymphoid cell lines engineered with inducible Cas9. We screened three ibrutinib-sensitive ABC lines, one ibrutinib-insensitive ABC line, and four ibrutinib-insensitive germinal centre B cell-like (GCB) lines, as well as two multiple myeloma and one T cell lymphoma line as controls (Extended Data Fig. 1a, Supplementary Tables 1, 2). For each gene, we derived a gene-level statistic that we term a CRISPR screen score (CSS), which is, in essence, the number of standard deviations away from the average effect of inactivating a gene (Supplementary Table 3, Methods).

Non-targeting control sgRNAs were not toxic, whereas sgRNAs targeting pan-essential genes[8] were depleted in all lines (Extended Data Fig. 1b). Among genes encoding B-cell transcription factors, we observed DLBCL subtype-specific dependencies in both GCB (MEF2B, TCF3, IRF8 and SPI1) and ABC (IRF4, SPIB and BATF) cell lines (Extended Data Fig. 1c). Results from a validation screen using approximately 10 sgRNAs per gene (Supplementary Tables 4, 5) were strongly correlated with those from the genome-wide screens ($P < 0.0001$; Supplementary Table 6, Extended Data Fig. 2a).

Most DLBCL lines depended on the BCR subunits CD79A and CD79B (Fig. 1), but engaged divergent downstream survival pathways. ABC lines uniquely relied on NF-κB regulators and on JAK1/STAT3 signalling triggered by the NF-κB-dependent cytokine IL-10. By contrast, BCR signalling in GCB lines was NF-κB-independent, but shared a dependence on PI3K/mTOR signalling with ABC cells, albeit using different signalling adapters (PIK3AP1 in ABC, CD19 in GCB). The BCR signalling mode in GCB is similar to that observed in another germinal-centre-derived malignancy, Burkitt lymphoma[9], which was previously known as tonic signalling because it resembled tonic, NF-κB-independent BCR signalling in naive mouse B cells[10]. However, GCB and Burkitt cell lines depended on both CD19 and LYN (Fig. 1, Extended Data Fig. 3), which are not required for tonic signalling in mouse B cells, so we instead term this phenomenon constitutive germinal centre BCR signalling.

The survival of BCR-dependent ABC lines relied on Toll-like receptor (TLR9), which coordinates MYD88 signalling in innate immune cells, and on two chaperones that regulate the subcellular localization of TLR9, CNPY3 and UNC93B1. TLR9 was the only essential TLR in ABC lines (Extended Data Fig. 2b). We validated these findings using time-dependent toxicity assays in 12 DLBCL lines transduced

[1]Lymphoid Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [2]Biometric Research Branch, Division of Cancer Diagnosis and Treatment, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [3]Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Gaithersburg, MD, USA. [4]Department of Medicine II, Hematology/Oncology, Goethe University, Frankfurt, Germany. [5]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [6]Departments of Pathology, City of Hope National Medical Center, Duarte, CA, USA. [7]British Columbia Cancer Agency, Vancouver, British Columbia, Canada. [8]Hospital Clinic, University of Barcelona, Barcelona, Spain. [9]Institute of Pathology, University of Würzburg, and Comprehensive Cancer Center Mainfranken, Würzburg, Germany. [10]Department of Clinical Pathology, Robert-Bosch-Krankenhaus, and Dr. Margarete Fischer-Bosch Institute for Clinical Pharmacology, Stuttgart, Germany. [11]University Health Network, Laboratory Medicine Program, Toronto General Hospital and University of Toronto, Toronto, Ontario, Canada. [12]Department of Laboratory Medicine and Pathology, Mayo Clinic, Scottsdale, AZ, USA. [13]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [14]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [15]Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [16]Experimental Pathology Laboratory, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [17]German Cancer Research Center and German Cancer Consortium, Heidelberg, Germany. [18]Present address: Aix-Marseille University, CNRS, INSERM, Centre d'Immunologie de Marseille-Luminy, Marseille, France. [19]These authors contributed equally: James D. Phelan, Ryan M. Young. [20]These authors jointly supervised this work: Thomas Oellerich, Louis M. Staudt. *e-mail: thomas.oellerich@kgu.de; lstaudt@mail.nih.gov

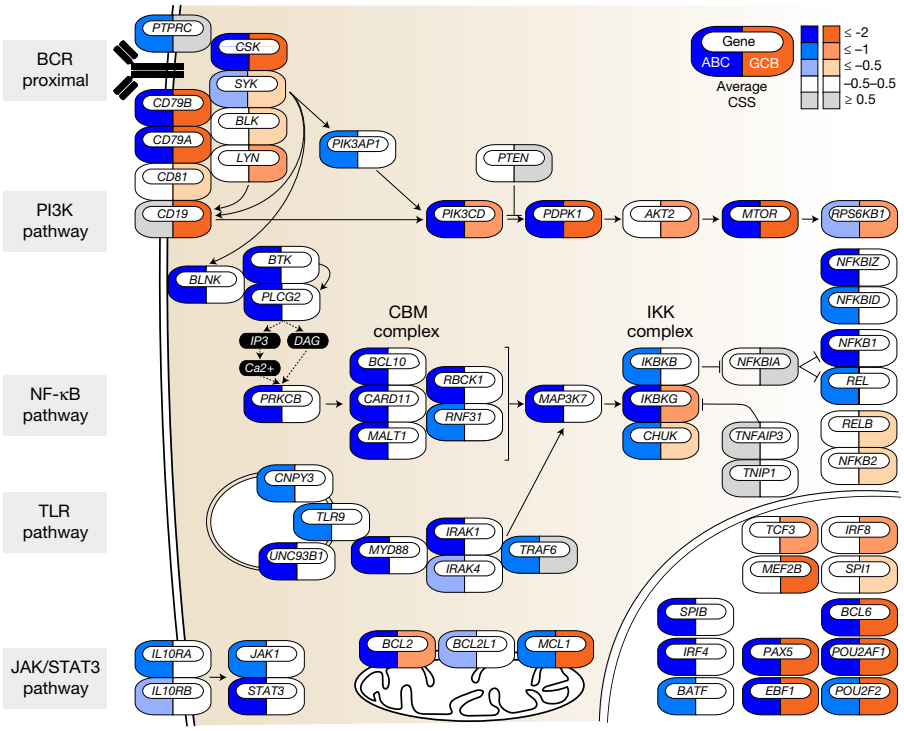


Fig. 1 | Genes essential for oncogenic signalling in lymphoma. Icons indicate essential genes from CRISPR screens coloured by the average CSS in GCB-dependent (orange) or BCR-dependent (blue) ABC DLBCL lines.

with vectors that co-express sgRNAs and green fluorescent protein (GFP) (Fig. 2a). As expected, ABC lines expressing mutant isoforms of MYD88 were sensitive to MYD88 deletion[7]. By contrast, TLR9, and its chaperones UNC93B1 and CNPY3[11], were only essential in ABC lines with *MYD88*^{L265P} and either a *CD79A* or *CD79B* mutation. These double-mutant lines were also particularly sensitive to BTK inhibition (Supplementary Table 1).

We next investigated copy number and gene expression levels of TLR pathway genes in 574 DLBCL tumours[12]. ABC tumours had recurrent single copy gains or amplifications involving *MYD88*, *TLR9*, *CNPY3* and *UNC93B1*, all of which were more highly expressed in ABC tumours and their expression correlated with copy number (Fig. 2b, Extended Data Fig. 4a, Supplementary Table 7). Altogether, 49.7% of ABC tumours had increased copy number of one or more of these genes (Fig. 2b, Supplementary Table 8), with *CNPY3* and *UNC93B1* demonstrating minimal common amplified regions of 1.1 Mb and 277 kb respectively (Extended Data Fig. 4b, Supplementary Table 9). These data provide genetic evidence that the TLR9 pathway contributes to the ABC phenotype.

To determine TLR9 function in ABC DLBCL, we expressed a fusion protein linking TLR9 to BioID2, a promiscuous biotin ligase that biotinylates proteins within approximately 10 nm[13]. Biotinylated proteins in TLR9–BioID2-expressing ABC cells were purified and compared to proteins from control cells by SILAC (stable isotope labelling by amino acids in cell culture)-based quantitative mass spectrometry. To define the TLR9 interactome that is essential in ABC DLBCL, we compared the enrichment of each protein quantified by mass spectrometry with its respective CSS metric (Fig. 2c). The TLR9-essential interactome

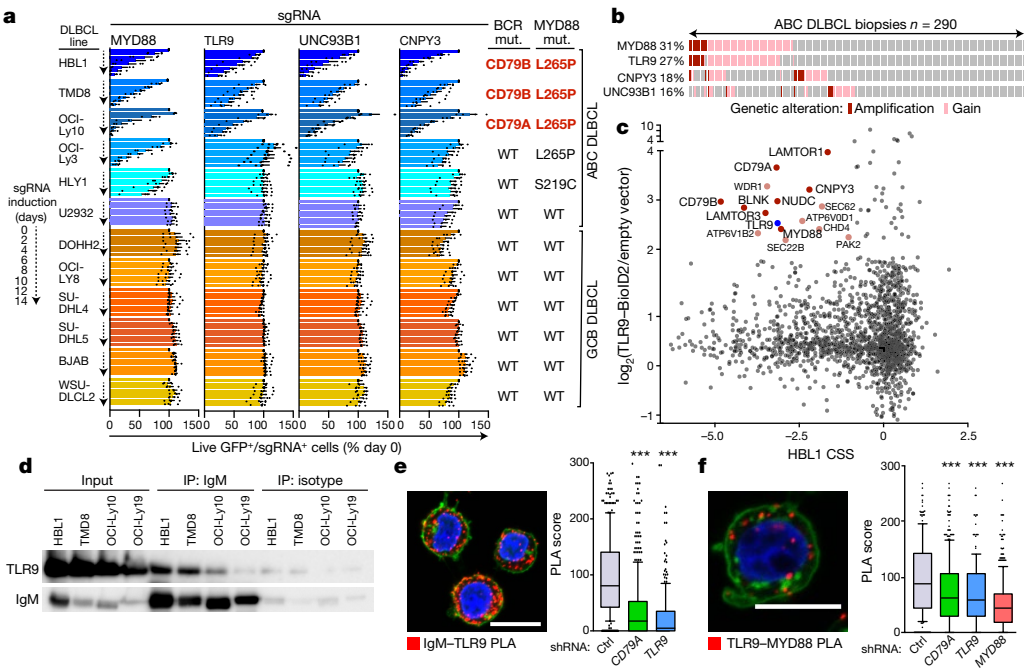


Fig. 2 | TLR9 couples BCR signalling and mutant MYD88. **a**, Toxicity of sgRNAs in DLBCL lines normalized to day 0. WT, wild type. **b**, Copy number gain or amplification of indicated genes in ABC biopsies. **c**, TLR9–BioID interactome in HBL1 cells versus CSS. Bait (TLR9) is labelled in blue; essential interactors are labelled red; essential interactors also in TMD8 cells are labelled dark red. **d**, TLR9 co-immunoprecipitates with IgM in ABC lines (HBL1, TMD8 and OCI-Ly10). OCI-Ly19 is a GCB line. **e**, **f**, Left, confocal images of PLAs (red) showing the IgM–TLR9 (**e**) or TLR9–MYD88 (**f**) interaction in HBL1 cells. Cells were counterstained with DAPI (blue) and wheat germ agglutinin (WGA; green). Right, PLA scores after knockdown of indicated genes. ***$P \leq 0.001$; see 'Statistics and reproducibility' section of the Methods for further information. Scale bar, 10 μm.

confirmed the association of TLR9 with MYD88 and CNPY3, but also revealed interactions with the BCR subunits CD79A and CD79B (Fig. 2c, Extended Data Fig. 4c–e, Supplementary Tables 10, 11). The IgM component of the endogenous BCR co-immunoprecipitated with TLR9 in three ABC lines more than in a GCB line (Fig. 2d). By contrast, neither TLR4 nor TLR7 co-immunoprecipitated with IgM (Extended Data Fig. 5a). TLR9 associated with IgM in an intracellular fraction of ABC cells rather than a plasma membrane fraction (Extended Data Fig. 5b), suggesting that the BCR and TLR9 might cooperate at an intracellular location.

To visualize where TLR9 and the BCR interact, we used proximity ligation assays (PLA), which identify proteins within tens of nanometres of each other[14]. An IgM–TLR9 PLA produced fluorescent puncta in the cytoplasm of ABC cells that was reduced by depletion of CD79A or TLR9 (Fig. 2e, Extended Data Fig. 5c). The IgM–TLR9 PLA signal was present across a panel of BCR-dependent ABC lines, with higher signals in double-mutant lines, whereas BCR-independent ABC and GCB lines had substantially lower signals (Extended Data Fig. 5d–f). IgG–TLR9 PLA gave no detectable signal (Extended Data Fig. 5g). IgM–TLR9 PLA signals co-localized with the endolysosomal marker LAMP1 (Extended Data Fig. 5h, i), consistent with the dependence of these ABC lines on UNC93B1 and CNPY3, which facilitate TLR9 entry into LAMP1$^+$ endolysosomes[11]. Ectopic expression of TLR9, wild-type MYD88 or the MYD88(L265P) mutant increased the IgM–TLR9 PLA signal (Extended Data Fig. 5j), suggesting that TLR9/MYD88 copy number gains in ABC tumours could augment BCR-TLR9 cooperation.

Knockdown of TLR9 decreased NF-κB-dependent gene expression and reduced IκB kinase activity in ABC lines with MYD88(L265P), confirming the role of TLR9 in oncogenic NF-κB signalling (Extended Data Fig. 6). TLR9–MYD88 PLA puncta were visible in the cytoplasm of ABC lines, but were diminished by the knockdown of TLR9, MYD88 or CD79A, suggesting that the BCR facilitates recruitment of MYD88 to TLR9 (Fig. 2f).

These results suggest that TLR9 coordinates signalling between the BCR and MYD88. We proposed that the BCR, TLR9 and MYD88 nucleate a signallosome that activates NF-κB, which we will term the MyD88–TLR9–BCR (My-T-BCR) supercomplex. To identify additional components of the My-T-BCR supercomplex, we expressed a MYD88(L265P)–BioID2 protein in three ABC lines and performed mass spectrometry analysis of MYD88-proximal biotinylated proteins. We identified proteins biotinylated in all three lines and used their CSS scores to define the essential MYD88 interactome, which included the BCR (CD79B), mTOR, PLCγ2 and the CARD11–BCL10–MALT1 (CBM) complex (Fig. 3a, Extended Data Fig. 7a, b, Supplementary Tables 12–14). Streptavidin pulldown and immunoblot analysis confirmed CARD11 and MALT1 biotinylation in ABC cells with MYD88(L265P)–BioID2 (Extended Data Fig. 7c, d).

Finding the CBM complex in proximity to MYD88 was unexpected since these adaptors are thought to independently promote NF-κB activation. Both MALT1–MYD88 and BCL10–MYD88 PLAs yielded robust cytoplasmic puncta in ABC cells, confirming the association of endogenous MYD88 with the CBM complex (Fig. 3b, Extended Data Fig. 7e, f). These PLA signals were reduced by knockdown of CD79A, TLR9 and CARD11, suggesting that BCR and TLR9 signalling cooperate to assemble MYD88 and the CBM into a supercomplex. Moreover, CARD11–BCL10 PLA puncta were reduced by the knockdown of TLR9 or MYD88 in double-mutant cell lines, demonstrating that TLR9 signalling controls CBM complex assembly in ABC cells (Fig. 3d, Extended Data Fig. 7g).

NF-κB is activated by IκB kinase (IKK)-dependent phosphorylation of IκBα. By PLA, both IgM and TLR9 associated with phosphorylated IκBα (p-IκBα) in the cytoplasm of ABC cells, which was reduced by knockdown of CD79A, TLR9 or MYD88 (Fig. 3d, e). Thus, NF-κB activation is closely associated with the My-T-BCR supercomplex.

We next visualized the subcellular location of the My-T-BCR supercomplex by staining ABC cells bearing MYD88(L265P)–BioID2 with

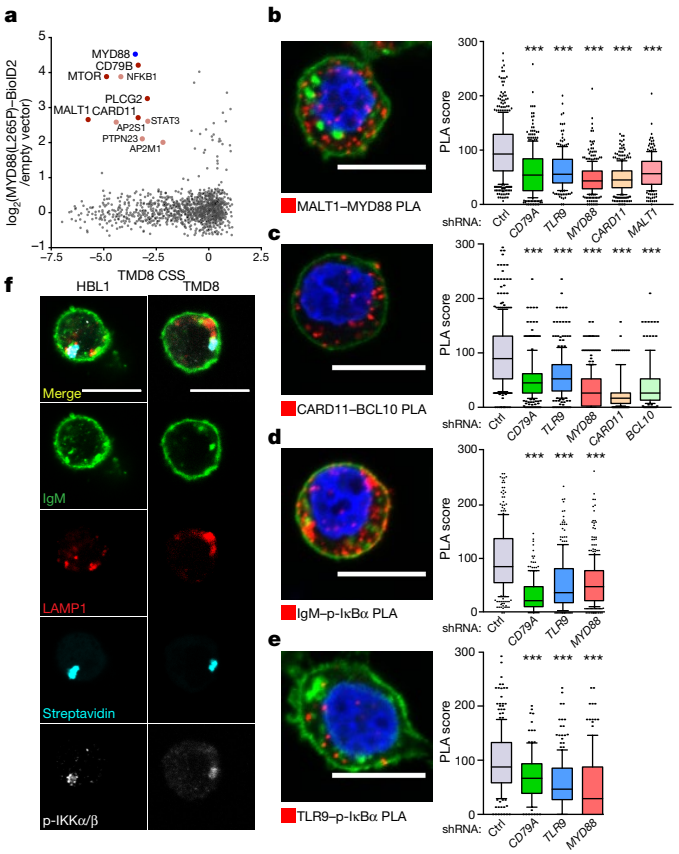


Fig. 3 | The My-T-BCR supercomplex coordinates NF-κB activation. **a**, MYD88(L265P)–BioID interactome in TMD8 cells versus CSS. Bait (MYD88 L265P) is labelled in blue; essential interactors are labelled red; essential interactors in at least two ABC lines are labelled dark red. **b**–**e**, Left, confocal image of PLAs (red) showing interaction of MALT1–MYD88 (**b**), CARD11–BCL10 (**c**), IgM–p-IκBα (**d**) and TLR9–p-IκBα (**e**). Cells were counterstained with DAPI (blue) and WGA (green). Right, PLA scores in HBL1 cells after shRNA knockdown of the indicated genes. ***$P \leq 0.001$; see Methods. **f**, Confocal images of MYD88(L265P)–BioID-transduced HBL1 or TMD8 cells stained as indicated. Scale bar, 10 μm.

fluorescently labelled streptavidin. The MYD88–BioID2 signal defined large cytoplasmic structures that co-localized with p-IKK, consistent with active NF-κB signalling at these sites (Fig. 3f, Supplementary Video 1). These complexes extended into the cytoplasmic space from the surface of LAMP1$^+$ vesicles. BCR was visualized by cell-surface labelling of IgM with a fluorescent Fab fragment on ice, followed by brief warming to allow internalization. The LAMP1$^+$ vesicles with MYD88–BioID2 signals also contained IgM, suggesting a dynamic shuttling of the BCR from the plasma membrane to the intracellular site of My-T-BCR supercomplex formation.

Given that the My-T-BCR supercomplex coordinates pro-survival signalling in ABC DLBCL, we hypothesized that inhibition of BTK activity by ibrutinib might disrupt this signalling complex. Ibrutinib reduced puncta of the My-T-BCR supercomplex in ABC lines bearing MYD88(L265P)–BioID2 (Extended Data Fig. 7h). To globally assess the effect of ibrutinib on the My-T-BCR supercomplex, we treated two ABC lines bearing MYD88(L265P)–BioID2 with ibrutinib and analysed the biotinylated proteins by mass spectrometry. Interactions of MYD88 with the CBM complex (CARD11), PLCγ2 and mTOR were disrupted by ibrutinib (Fig. 4a, Extended Data Fig. 7i and Supplementary Tables 14, 15).

The ibrutinib-sensitive association of mTOR with MYD88 suggested that signalling by the My-T-BCR supercomplex might affect pathways controlled by mTOR. Of note, components of the Ragulator complex (LAMTOR1, LAMTOR3, LAMTOR4 and RRAGA), which regulates mTORC1 activity at the lysosomal membrane[15], were biotinylated

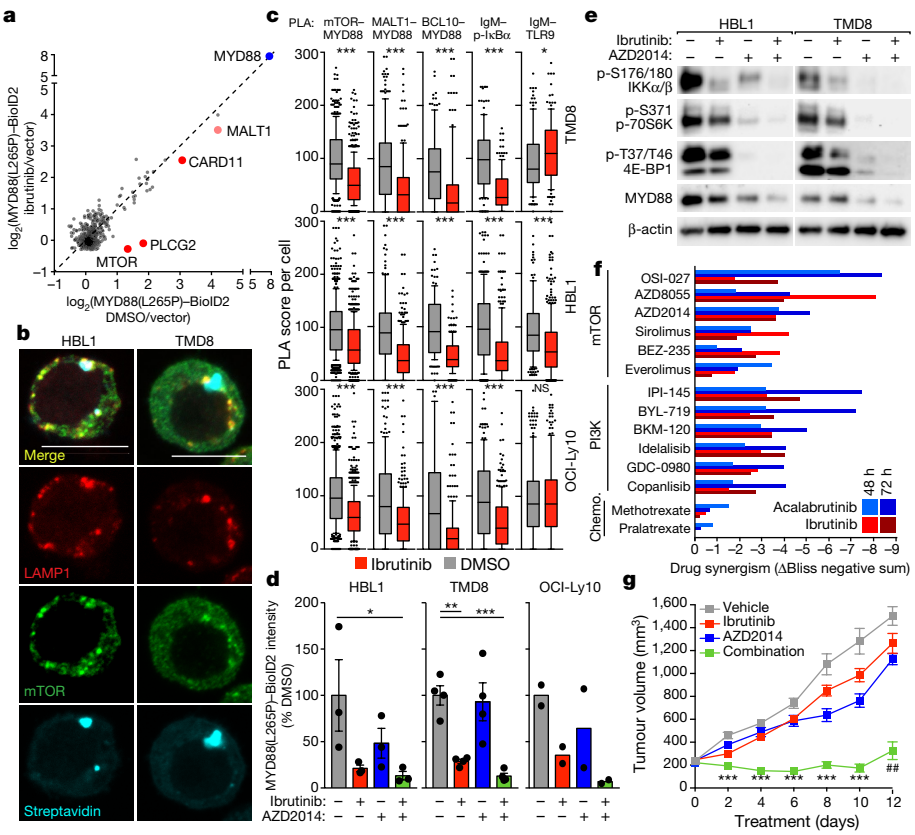


Fig. 4 | mTOR is an essential component of the My-T-BCR supercomplex. **a**, MYD88(L265P)–BioID interactome in OCI-Ly10 cells treated with ibrutinib (10 nM) or DMSO. Ibrutinib-sensitive interactions are labelled in red. Bait (MYD88(L265P)) is labelled in blue. **b**, Confocal images of mTOR (green), LAMP1 (red) and MYD88(L265P)–BioID2 (cyan, streptavidin) in ABC cells. Scale bar, 10 μm. **c**, PLA scores for indicated protein interactions in ABC lines treated with ibrutinib or DMSO. **d**, Normalized MYD88(L265P)–BioID intensity per cell in ABC lines treated as indicated for 24 h. **e**, Immunoblots using the indicated antibodies of ABC lines treated with indicated drugs for 24 h. **f**, Synergistic toxicity scores in TMD8 cells treated with ibrutinib or acalabrutinib together with the indicated drugs. **g**, Growth of TMD8 xenografts in NSG mice treated as indicated. NS, not significant; $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; see Methods. Double hash symbols (##) denote mouse death.

by TLR9–BioID2, as were components of the lysosomal V-ATPase (ATP6V1B2 and ATP6V0D1), which regulates the mTORC1 response to amino acids[16] (Fig. 2c, Extended Data Fig. 4c, Supplementary Tables 10, 11). In ABC lines with MYD88(L265P)–BioID2, mTOR localized to LAMP1$^+$ vesicles, often in proximity to the My-T-BCR supercomplex (Fig. 4b). PLA in three ABC lines confirmed that ibrutinib decreased association of endogenous MYD88 with mTOR, MALT1 and BCL10 (Fig. 4c). Ibrutinib also decreased the association of IgM and p-IκBα, but had mixed effects on IgM association with TLR9. These findings suggest that IgM trafficking to TLR9$^+$ endolysosomes is constitutive, but interaction of the My-T-BCR supercomplex with mTOR, the CBM complex, and NF-κB is controlled by BTK-dependent BCR signalling.

Given the proximity of MYD88 and mTOR, we investigated the effect of mTOR inhibition on the My-T-BCR supercomplex. In MYD88(L265P)–BioID2-expressing ABC cells, formation of the

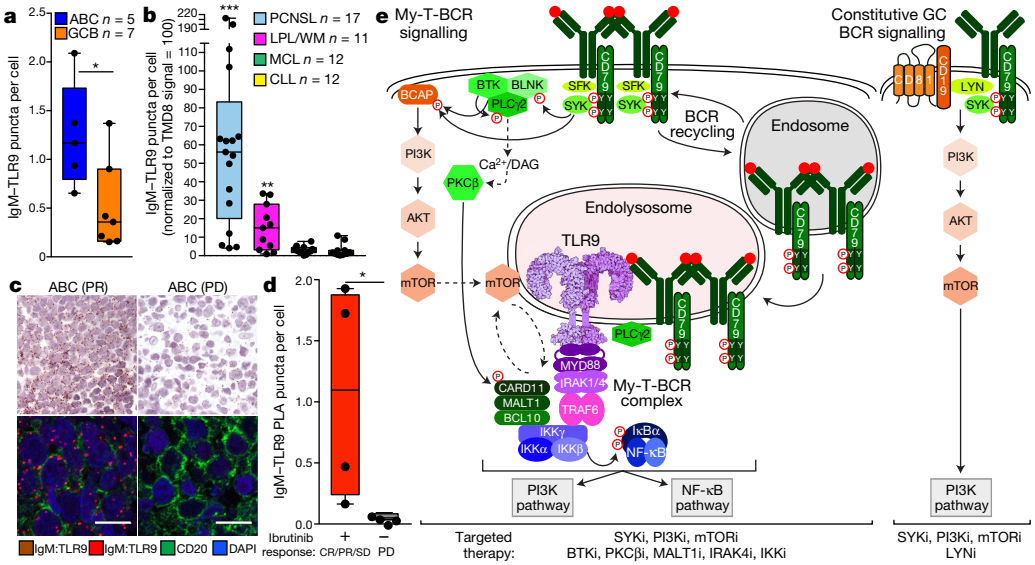


Fig. 5 | The My-T-BCR supercomplex identifies ibrutinib-responsive lymphomas. a, **b**, IgM–TLR9 PLA puncta per cell in GCB and ABC biopsies (**a**) or indicated lymphoma biopsies (normalized to TMD8 signal = 100) (**b**). CLL, chronic lymphocytic leukaemia; LPL, lymphoplasmacytic lymphoma; MCL, mantle cell lymphoma; PCNSL, primary central nervous system lymphoma; WM, Waldenström macroglobulinaemia. PLA data normalized to TMD8 control **c**, Representative IgM–TLR9 PLA images of ABC biopsies. Top, bright-field images; bottom, fluorescence images. Scale bar, 10 μm. PD, progressive disease; PR, partial response. **d**, IgM–TLR9 PLA of biopsies from DLBCL patients treated with ibrutinib. Red denotes responders (red; complete response (CR)/PR/stable disease (SD)); grey denotes non-responders (PD). **e**, Models of My-T-BCR supercomplex signalling and constitutive germinal centre (GC) BCR signalling. $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; see Methods.

My-T-BCR supercomplex was reduced by ibrutinib, but was further attenuated by the addition of AZD2014, an mTORC1/2 inhibitor (Fig. 4d). Dual mTOR and BTK inhibition cooperatively decreased MYD88 protein levels and blocked mTOR activity, as assessed by p-4E-BP1 and p-S6 kinase, as well as NF-κB activation, as assessed by p-IKK (Fig. 4e). These data provide mechanistic insights into the synergism between BTK inhibitors and drugs targeting mTOR or PI3K in ABC models growing in vitro (Fig. 4f) and in vivo[17,18] (Fig. 4g).

Finally, we examined whether the My-T-BCR supercomplex is detectable in primary lymphoma biopsy samples, and if its presence might be associated with ibrutinib responsiveness. We optimized the PLA for use in formalin-fixed biopsy samples using a tissue microarray of 81 lymphoma cell lines. IgM–TLR9 PLA signals were highest in ABC lines with chronic active BCR signalling, with little if any PLA signal in other lymphoma lines or normal B cells present in tonsils or reactive lymph nodes (Extended Data Fig. 8a, b, Supplementary Table 17). Among DLBCL biopsies, ABC cases had significantly more IgM–TLR9 puncta than GCB cases (Fig. 5a). High IgM–TLR9 PLA signals were also observed in the biopsies of primary central nervous system lymphoma, Waldenström macroglobulinaemia, and its relative, lymphoplasmacytic lymphoma (Fig. 5b). These malignancies commonly have *MYD88*^{L265P} and/or *CD79A* or *CD79B* mutations, and respond frequently to ibrutinib[19–23]. Of two Waldenström macroglobulinaemia lines tested, one had My-T-BCR supercomplexes, and knockdown of the BCR (CD79A) or TLR9 was selectively toxic for this line (Extended Data Fig. 9a–c). My-T-BCR supercomplexes were not evident in mantle cell lymphoma or chronic lymphocytic leukaemia samples (Fig. 5b), suggesting that these malignancies rely on a qualitatively distinct form of BCR signalling.

We next examined eight available biopsies from patients with relapsed or refractory DLBCL enrolled on a clinical trial of ibrutinib monotherapy[1]. We adapted the IgM–TLR9 PLA to allow immunohistochemical identification of CD20^{+} lymphoma cells (Fig. 5c). Three ABC cases and one unclassified DLBCL scored positive in the IgM–TLR9 PLA while three other ABC cases and one GCB case were negative (see Methods; Supplementary Table 16). The percentage of IgM–TLR9 PLA-positive malignant cells was significantly higher ($P = 0.0286$) in tumours that responded to ibrutinib than in those that progressed on treatment (Fig. 5d). In this series, two responding cases with IgM–TLR9 puncta had *CD79B* or *CD79A* mutations, respectively, but lacked MYD88(L265P), while two other responders were wild-type for these genes (Supplementary Table 16). These findings demonstrate that the My-T-BCR supercomplex exists in ABC DLBCL tumours that respond to ibrutinib, even in those lacking the *MYD88*^{L265P}*CD79A* or *MYD88*^{L265P}*CD79B* double-mutant genotype.

We provide genetic, proteomic, cell biological and functional evidence for a pro-survival signalling hub—termed the My-T-BCR supercomplex—that coordinates NF-κB activation in DLBCL and identifies tumours that respond to therapeutic inhibition of NF-κB by ibrutinib. This supercomplex is present in a subset of ABC DLBCL lines and tumours, but is generally absent from GCB DLBCL, which have an alternative 'constitutive germinal centre' BCR signalling mode, requiring distinct therapeutic strategies (Fig. 5e). The My-T-BCR supercomplex provides mechanistic insight into the efficacy of drug combinations in ABC DLBCL and may aid in the development of predictive assays to identify patients who would benefit from drugs that target BCR-dependent NF-κB activation, including BTK inhibitors.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0290-0.

Received: 12 October 2017; Accepted: 13 June 2018;
Published online 20 June 2018.

1. Wilson, W. H. et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat. Med.* **21**, 922–926 (2015).
2. Alizadeh, A. A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
3. Wright, G. et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl Acad. Sci. USA* **100**, 9991–9996 (2003).
4. Young, R. M. et al. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. *Proc. Natl Acad. Sci. USA* **112**, 13447–13454 (2015).
5. Davis, R. E. et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88–92 (2010).
6. Lenz, G. et al. Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676–1679 (2008).
7. Ngo, V. N. et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
8. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
9. Schmitz, R. et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116–120 (2012).
10. Lam, K. P., Kühn, R. & Rajewsky, K. In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* **90**, 1073–1083 (1997).
11. Saitoh, S. & Miyake, K. Regulatory molecules required for nucleotide-sensing Toll-like receptors. *Immunol. Rev.* **227**, 32–43 (2009).
12. Schmitz, R. et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N. Engl. J. Med.* **378**, 1396–1407 (2018).
13. Kim, D. I. et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196 (2016).
14. Söderberg, O. et al. Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay. *Methods* **45**, 227–232 (2008).
15. Sancak, Y. et al. Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* **141**, 290–303 (2010).
16. Zoncu, R. et al. mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H^{+}-ATPase. *Science* **334**, 678–683 (2011).
17. Mathews Griner, L. A. et al. High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proc. Natl Acad. Sci. USA* **111**, 2349–2354 (2014).
18. Ezell, S. A. et al. Synergistic induction of apoptosis by combination of BTK and dual mTORC1/2 inhibitors in diffuse large B cell lymphoma. *Oncotarget* **5**, 4990–5001 (2014).
19. Grommes, C. et al. Ibrutinib unmasks critical role of bruton tyrosine kinase in primary CNS lymphoma. *Cancer Discov.* **7**, 1018–1029 (2017).
20. Lionakis, M. S. et al. Inhibition of B cell receptor signaling by ibrutinib in primary CNS lymphoma. *Cancer Cell* **31**, 833–843 (2017).
21. Nakamura, T. et al. Recurrent mutations of CD79B and MYD88 are the hallmark of primary central nervous system lymphomas. *Neuropathol. Appl. Neurobiol.* (2015).
22. Treon, S. P. et al. Ibrutinib in previously treated Waldenström's macroglobulinemia. *N. Engl. J. Med.* **372**, 1430–1440 (2015).
23. Treon, S. P. et al. MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N. Engl. J. Med.* **367**, 826–833 (2012).

Acknowledgements This research was supported by the Intramural Research Program of the NIH, CCR, NCI and by a NCI Strategic Partnering to Evaluate Cancer Signatures (SPECS II) grant (5U01CA157581-05), as well as by Deutsche Krebshilfe (#111399) and Deutsche Forschungsgemeinschaft (SFB1177). D.E.W. is a Damon Runyon Fellow (DRG-2208-14). S.R. is a H2020 Marie Sklodowska-Curie global fellow (#661066). We thank AstraZeneca for AZD2014.

Reviewer information *Nature* thanks A. Alizadeh, I. Mellinghoff, A. Rothstein and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.D.P. and R.M.Y. designed and performed experiments, analysed data, made the figures and wrote the manuscript. D.E.W., M.J.K., S.P., S.R., M.K., A.L.S.III, M.C., J.Q.W., R.S., M.N. and E.B. designed and performed experiments and analysed data. G.W.W., D.W.H., L.C. and C.J.T. analysed data, Y.J., Y.Y., H.Z., X.Y., W.X., M.M.P., R.R.V. and T.D.-H. performed experiments, W.H.W., W.C.C., E.S.J., R.D.G., E.C., A.R., G.O., J.D., L.M.R., F.J.R., F.E., M.H. and S.M.H. provided annotated clinical samples. T.O. and L.M.S. designed experiments, analysed data, made the figures and wrote the manuscript. The authors state they have no competing interests.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0290-0.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0290-0.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.O. or L.M.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Cell culture. Cell lines were grown at 37 °C in the presence of 5% CO_2 and maintained in RPMI supplemented with fetal bovine serum (Tet tested, Atlanta Biologics,) and 1% penicillin/streptomycin and 1% L-glutamine (Invitrogen), except for OCI-Ly10 and OCI-Ly3 which were grown in IMDM supplemented with 20% heparinized human plasma, 1% penicillin/streptomycin and 55 μM β-mercaptoethanol. All cell lines were regularly tested for mycoplasma using the MycoAlert Mycoplasma Detection Kit (Lonza) and DNA fingerprinted by examining 16 regions of copy number variants (Jonathan Keats, personal communication). OCI-Ly3, although present in the database of commonly misidentified cell lines maintained by ICLAC, was included in this study as a necessary model of a BCR-independent, *MYD88*^{L265P} mutant, MYD88-dependent ABC DLBCL. This cell line was authenticated by DNA fingerprinting and compared to historical DNA controls.

Cas9 vector construction. pRetroCMV/TO-Cas9-Hygro was created by ligating the tetracycline-inducible CMV promoter from pcDNA4/TO (Invitrogen) with MfeI/XbaI and blunt cloned into the XhoI/EcoRI digested pRetrosuper vector[24]. The puromycin-resistance gene from pRetroCMV/TO was removed with StuI/ClaI and replaced with PGK-hygromycin, which was isolated from pMSCV Hygro (Clonetech) with AgeI/HindIII and similarly cloned into pRetroCMV/TO. Cas9 was isolated from the LentiCrispr v2 (Addgene 52961) and blunt cloned into pRetroCMV/TO-hygro digested with AgeI/BamHI. pCW-Cas9-Blasticidin was generated from pCW-Cas9-puro which was purchased from Addgene (50661) and digested with BamHI and XbaI to remove the puromycin resistance gene. A g-block (IDT) containing the blasticidin resistance gene was Gibson cloned into the cut vector with 12-base-pair overlaps.

Cas9 clone generation. Cell lines were transduced several times with either pTO-Cas9-hygro or pCW-Cas9, selected and dilution cloned. Single-cell clones were picked and tested for functional Cas9 cutting after transduction with sgRNAs that target surface markers including CD20 or ICAM1. Clones were selected based on loss of surface expression within the transduced population as measured by FACS 8–14 days after the addition of doxycyline.

sgRNA vector and cloning. The pLKO-based sgRNA vector was purchased Addgene (52628). The puromycin gene was removed and replaced with a puro-GFP fusion protein previously described[25] using Gibson assembly. The resulting plasmid was digested with BfuAI and incubated with shrimp alkaline phosphatase before isolating the backbone. Complementary sgRNA sequences flanked by ACCG on the 5′ end, and CTTT on the 3′ of the reverse strand, were annealed, diluted and ligated into the cut vector with T4 ligase according to the manufacturer's instructions. All transformations were performed in *Stbl3* bacteria and grown at 30 °C.

sgRNA library construction. The genome-wide Brunello sgRNA library[26] was purchased from Addgene and transformed in *Stbl4* bacteria from Invitrogen. The Brunello library contains 77,441 sgRNAs targeting four unique positions in most (19,114) protein-coding genes, along with 1,000 negative control sgRNAs. Sequences for the follow up library of 12,472 sgRNAs were chosen from published sgRNA libraries[27,28] or were designed using the online tool at http://crispr.mit.edu. The library (CustomArray Inc.) of 74-mer of the sgRNA sequence prepended with the oligonucleotide sequence GGAAAGGACGAAACACCG and followed by GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC. The oligonucleotide library was PCR amplified with Herculase II Fusion DNA polymerase (Agilent) using ArrayF and ArrayR[27]. The subsequent PCR product was gel extracted using an eGel from Invitrogen and 20 ng of library was Gibson cloned into BfuAI-cut sgRNA vector following the manufacturer's instructions. Transformations were grown at 30 °C overnight on 24.5 cm^2 bioassay plates maintaining at least 100× coverage. Colonies were scrapped, spun and DNA was isolated with Blood and Cell Culture DNA Maxi kits (Qiagen).

Virus production and transduction. Lentiviruses were produced in 293FT cells by transfecting sgRNA vectors with packaging vectors pPAX2 (Addgene 12260) and pMD2.g (Addgene 12259) in a 4:3:1 ratio in serum-free Opti-MEM. Trans-IT 293T (Mirus) was added and incubated for 15 min before adding dropwise to cells. Supernatants were harvested 24, 48 and 72 h later, spun at 1,000 *g* to pellet any virus producing cells and then incubated with Lenti-X concentrator (CloneTech). Virus was concentrated according to manufacturer's instructions, aliquoted and frozen. Virus titration was performed on target cell populations and GFP was measured 3–4 days later. When GFP was not present in the backbone of the sgRNA plasmids, transduced cells were split and incubated with or without puromycin until untransduced control cells were dead. The percentage of viable cells was then measured by FACS and percent transduction was calculated as the ratio of viable cells in treated versus untreated wells.

Pooled sgRNA screening. For both genome-wide and targeted follow-up screens, individual replicates were transduced such that an average of 500 copies of each sgRNA was present after selection. Cultures were carried for the duration of the 21-day screen maintaining 500× coverage. Antibiotic selection was started 3–4 days after transduction and carried out until untransduced control cells were dead, approximately 4–5 days later. Cells were then harvested for a day 0 time point and doxycycline was added to the culture media at 200 ng ml^{-1} final concentration. Transduced cells were counted and passaged every two days with fresh media containing doxycycline until day 21 when cells were again collected for DNA extraction. DNA was isolated from frozen cell pellets using Qiagen QIAmp DNA Blood Midi and Maxi kits.

Library amplification, sequence extraction and PCR primers. For both screens, sgRNA sequences were amplified using a nested PCR to first isolate the sgRNA sequence from genomic DNA and then to add nextgen sequencing adapters compatible with Illumina's NextSeq500. Products were amplified using ExTaq (Takara) for 18 cycles in both rounds of amplification. Products were size selected using an eGel (Invitrogen) and libraries were quantitated using an Illumina specific Kapa quantification kit according to the manufacturer's instructions or by Qubit (Thermo Fisher Scientific). All libraries were sequenced using a high output single-read 75 cycle read flow cell. An average of 400× (200–700×) sequencing depth was achieved. Libraries were multiplexed using indexes compatible with the Illumina TruSeq HT kit with the primers below, in which 'x' denotes an 8-base-pair index and 'y' represents a variable length adaptor inserted to prevent monotemplate. In total, 8 forward primers and 12 reverse primers were used following this format, such that 96 samples could be multiplexed. BaseSpace sequence hub (https://www.basespace.illumina.com/home/index) was used to evaluate sequencing quality measures and to demultiplex sequencing reads. Sequences were aligned to the sgRNAs library allowing for a one basepair mismatch using custom scripts and Bowtie 2 version 2.2.9 with the following parameters: -p 16 -f–local -k 10–very-sensitive-local -L 9 -N 1.

CRISPR screen primers. First PCR forward primer: AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG; first PCR reverse primer: GTAATTCTTTAGTTTGTATGTCTGTTGCTATTATG; second PCR forward primer: AATGATACGGCGACCACCGAGATCTACACxACACTCTTTCCCTACACGACGCTCTTCCGATCTyTCTTGTGGAAAGGACGAAACACCG; second PCR reverse primer: CAAGCAGAAGACGGCATACGAGATxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCtctactattctttcccctgcactgt. PCR amplification for sgRNA library construction: ARRAY-F: TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG; ARRAY-R: ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC.

Pooled sgRNA screen analysis. CSSs were calculated using the following formulas:

Step 1. Normalize raw counts by total read counts

$$N_{irjd} = 1 + (X^{\star}_{irjd} 4 \times 10^7) / \sum_i X_{irjd}$$

Step 2. Eliminate sgRNAs with low counts across all experimental conditions by calculating

$$m_i = \max_{jd}(\min_r N_{irjd})$$

and eliminating those sgRNAs for which $m_i < 100$.

Step 3. Calculate log ratios (LR)

$$\mathrm{LR}_{irj} = \log_2(N_{irj21}/N_{irj0})$$

Step 4. *Z*-transform log ratios

$$Z_{irj} = (\mathrm{LR}_{irj} - \mathrm{mean}_i(\mathrm{LR}_{irj}))/\mathrm{standard\ deviation}_i(\mathrm{LR}_{irj})$$

Step 5. Average across replicates

$$S_{ij} = \mathrm{mean}_r(Z_{ijr})$$

Step 5. Calculate signal variance (SV) for each sgRNA, using total variance (TV) and error variance (EV)

$$\mathrm{TV}_i = \mathrm{var}_j(S_{ij})$$
$$\mathrm{EV}_i = \mathrm{mean}_j(\mathrm{var}_r(Z_{irj})/R_i)$$
$$\mathrm{SV}_i = \mathrm{TV}_i - \mathrm{EV}_i$$

Step 6. For each gene *g*, let G_g be the set of sgRNAs that represent it, and calculate the maximal pair-wise correlation between any two sgRNAs in this set.

$$C_g = \max_{k,l \in G_g}(\mathrm{correlation}_j(S_{kj}, S_{lj}))$$

If for a given gene $C_g < 0.45$ then let i_g be the index of the sgRNA representing that gene that has the highest signal variance and use that as the sole representative of gene j $\text{Score}_{gj} = S_{i_g j}$

If $C_g \geq 0.45$, then proceed to steps 7 and 8.

Step 7. Average the two sgRNAs that were most correlated within gene g.

$$M_{gj} = \frac{1}{2}(S_{kj} + S_{lj})$$

in which $k,l \in G_g$ are such that

$$\underset{j}{\text{Correlation}}(S_{kj}, S_{lj}) = C_g$$

Step 8. For each $i \in G_g$ calculate

$$V_i = \underset{j}{\text{correlation}}(S_{ij}, M_{gj})/C_g$$

Average together those sgRNAs for which $V_i > 0.85$ to arrive at final CSS for gene g

$$\text{CSS}_{gj} = \underset{i \in G_g, V_{i>0.85}}{\text{mean}}(S_{ij})$$

In which i denotes [1–77,441] index indicating sgRNA; j denotes [1–11] index indicating cell line; R_j denotes the number of replicates for cell line j; r denotes the [1–R_j] index indicating replicate; d denotes the [0, 21] index indicating time point; X_{irjd} indicates the raw sequencing counts for sgRNA i, in replicate r of cell line j at time d.

For the replication sgRNA library, in which most genes had 9–10 sgRNAs per gene, we found this could be simplified by using the z-scores of the averaged $\log_2$ fold change of all sgRNAs per gene. As described above, using the best correlated sgRNAs per gene excluded many poor performing sgRNAs, it also excluded many high performing sgRNAs that shared expected subtype-specific toxicities.

Statistical significance. The statistical significance in Fig. 1 of the CSS of ABC BCR-dependent, or GCB DLBCL, compared to all other cell lines, was calculated with a two sided random variance t-test[29] for individual genes. Screen correlations were calculated using a Pearson correlation on gene-level metrics in GraphPad Prism 7.0 software on genes displayed in Extended Data Fig. 2.

FACS analysis. Cell lines were transduced with sgRNA vectors marked by GFP. Three to four days after transduction, GFP levels were measured by flow cytometry on a BD FACS Calibur using CellQuest Pro version 6.0 and analysed with FlowJo version 9. Cells were split every other day into doxycycline containing media and GFP levels were followed for 14 days and normalized to the day 0 measurement. All sgRNA and shRNA sequences are listed below. When surface proteins were targeted, knockout was validated by flow cytometry by spinning cells down, washing in FACS buffer (PBS plus 2% (v/v) FBS, 1 mM EDTA), and stained at 4 °C for 30 min in FACS buffer with fluorescently labelled antibodies: mouse anti-human CD19-APC (Biolegend SJ25C1, 1:500), mouse anti-human CD81-PE (Biolegend 5A6, 1:500), mouse anti-human IgM-APC (MHM-88), 1:400); or from Southern Biotech: goat anti-human IgG-PE (1:200).

Drug sensitivity assays. DLBCL cell lines were enumerated and 10,000 cells were seeded in triplicate in a 96-well plate in fresh media. Ibrutinib (Selleck Chem) was dissolved in DMSO and equal volumes of diluted drug were added to cells to reach the indicated final concentration. Cells were cultured with drugs, which were replenished after 48 h. Metabolic activity was measured at day 4 by adding 10 µl of MTS reagent (Promega) and incubating at 37 °C for 4 h. Absorbance was measured at 490 nm using a 96-well Tecan Infinite 200 Pro plate reader. Absorbance values from media-only wells were subtracted and data were normalized to DMSO control unless otherwise stated. GR50 was calculated using the online tool GRcalculator (http://www.grcalculator.org/)[30]. Drug matrix screens and ΔBliss calculations were performed as previously described[17].

Gene expression profiling and signature enrichment. Cells were transduced with shRNAs, puro-selected and collected at indicated times after shRNA induction. RNA was isolated using RNEasy mini kits (Qiagen). Gene expression was assessed using two-colour human Agilent 4 × 44 K gene expression arrays following the manufacturers protocol. In brief, control shSC4 (control, Cy3-labelled) RNA was compared to RNA from cells with shRNAs targeting *TLR9* (C4), TLR9 (D7), *MyD88* (A7) or *MyD88* (B3) (Cy5-labelled) at each of the indicated time points. Array elements were filtered for spot quality using Agilent Feature Extraction software version 10.7, specific genes were determined to be downregulated if the $\log_2$ fold change (comparing control shSC4 to shRNA for *TLR9*) was less than −0.3 for at least three of the four time points (shTLR9) per cell line. Signature enrichment was performed as previously described[31]. In brief, downregulated genes were tested for overlap with published gene signatures in a 2 × 2 contingency table using a Fisher's exact test.

DNA copy number analysis. DLBCL DNA samples were analysed with the Affymetrix SNP6.0 array. Probe log ratios were calculated using Affymetrix Genotyping Console, and were collected into segments of similar value using circular binary segmentation (https://bioconductor.org/packages/release/bioc/html/DNAcopy.html). These segments were assigned copy number values as previously described[32] without segment length restrictions. DNA copy number was correlated to sample gene expression using linear regressions calculated with Graphpad Prism version 7.0. Amplified regions were identified and visualized on the UCSC genome browser, Hg19.

NF-κB reporters. The generation of the IκBα luciferase reporter cell line has been previously described[33]. In brief, the TMD8-IκBα cell line was transduced with indicated shRNAs, puro-selected and induced with doxycycline. Cells were collected at the indicated time points and luciferase was measured with the Dual Luciferase Reporter Assay System (Promega) on a Tecan Infinite 200 Pro plate reader.

IgM co-immunoprecipitation. HBL1, TMD8, OCI-Ly10 and OCI-Ly19 cells were lysed at 10^7 cells per ml in a modified RIPA buffer (0.5% Triton X-100, 0.25% deoxycholate, 0.025% SDS, 10 mM Tris, pH 8.0, 100 mM NaCl, 10 mM EDTA, 1 mM Na_3VO_4, 30 mM pyrophosphate, 10 mM glycerophosphate, 1 mM AEBSF, 0.02 U ml^{-1} aprotinin and 0.01% NaN_3) for 10 min on ice. Lysates were cleared by centrifugation at 14,000 *g* for 20 min at 4 °C. IgM was immunoprecipitated by incubating lysates on ice for 1 h with 10 µg of biotin-labelled goat anti-human IgM (Jackson Immunoresearch), followed by the addition of 35 µl of pre-washed streptavidin-agarose beads (Invitrogen) and rotated for 30 min at 4 °C. Beads were washed three times with cold 1× RIPA buffer, then solubilized by adding 2× LDS sample buffer (Invitrogen) with 1% β-mercaptoethanol and boiled for 5 min. Samples were separated on a 10% polyacrylamide gel and transferred to Immobilon-p PVDF membrane (Millipore) for western blot analysis. Membranes were probed with rabbit anti-TLR9 monoclonal XP, rabbit anti-TLR7 (Cell Signaling Technologies), rabbit anti-TLR4 (Santa Cruz Biotechnology) and goat anti-IgM-HRP (Bethyl).

PLA. DLBCL cell lines were left untreated, treated with 10 nM ibrutinib, 200 nM AZD2014 or equivalent volumes of DMSO, or transduced with control shRNA (SC4) or shRNAs targeting *CD79A*, *TLR9*, *MYD88*, *CARD11*, *BCL10* or *MALT1*, followed by puromycin (Invitrogen) selection as previously described[34]. Cells were plated onto a 15 well µ-Slide Angiogenesis ibiTreat chamber slide (Ibidi) and allowed to adhere to the surface for 30 min at 37 °C. Cells were then fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 20 min at room temperature and then washed in PBS (Invitrogen). Cellular membranes were labelled with 5 µg ml^{-1} WGA conjugated to Alexa Fluor 488 (Thermo Fisher Scientific) for 10 min at room temperature. Cells were permeabilized in cold methanol for 10 min, washed in PBS and then blocked in Duolink Blocking buffer (Sigma) for 30 min at room temperature. Primary antibodies were diluted in Duolink Antibody Diluent (Sigma) and incubated overnight at 4 °C (see Supplementary Table 8). Where appropriate, cells were counterstained with mouse anti-LAMP1 conjugated to Alexa Fluor 405 (Santa Cruz Biotechnology) with the primary antibodies. The next morning, cells were washed for 20 min in a large volume of PBS with 1% BSA, followed by addition of the appropriate Duolink secondary antibodies (Sigma), diluted and mixed according to the manufacturer's instructions. Cells were incubated for 1 h at 37 °C, after which cells were washed in TBST with 0.5% Tween-20 for 10 min. Ligation and amplification steps of the PLA were performed using the Duolink in situ Detection Reagents Orange kit (Sigma) according to the manufacturer's instructions. Following the PLA, cells were mounted in Prolong Gold mounting media with DAPI (Invitrogen). Images were acquired on a Zeiss LSM 880 Confocal microscope using Zeiss Zen Black version 2.3. Images for display and Pearson's correlation coefficients values were calculated with NIH ImageJ/FIJI software version 2.0.0-rc-65/1.5ls[35]. PLA spots were counted in cell lines using Blobfinder version 3.2[36]. PLA scores were determined by normalizing the number of PLA spots counted in each sample to the average number of PLA spots counted in the control sample, which was set to 100. Box and whisker plots display the median PLA score with whiskers incorporating 10–90% of all data, outliers are displayed as dots.

The PLA was performed on formalin-fixed, paraffin-embedded (FFPE) tissue microarrays or biopsy samples in a similar manner. FFPE microarrays (7 µm) and patient tissue sections (4 µm) were deparaffinized in xylene and rehydrated in graded alcohol and distilled water. Heat-induced antigen retrieval was performed on tissue microarrays and tissue sections at pH 6.0 for 30 min. Slides were then placed in Tris-buffered solution and prepared for proximity ligation assay, as described above, samples were costained with mouse anti-human CD20-eFluor660 or AlexaFluor488 (L26, eBioscience). Data were analysed using Blobfinder version 3.2. Cells with 10 or more IgM–TLR9 puncta in their nucleus were removed from analysis to control for increased autofluorescence in FFPE samples. Tissue microarrays were prepared by fixing cells in neutral buffer formalin for 24 h, pelleting and resuspending in an equal volume of low-melt agarose in a 10 ml conical tube. The resulting pellet was paraffin embedded by standard protocol[37]. The resultant blocks were used to construct a cell line array (CMA) using the same approach

used for construction of a tissue microarray, with 1.00 mm needles, using a Beecher MTA-1 instrument (Beecher Instruments). Sample identifiers were removed and blinded before pathology review for PLA signal. After all data were collected, sample identifiers were revealed and samples were grouped by response to ibrutinib.

BioID2 contructs. BioID2 (Addgene 80899) was appended to the C terminus of TLR9 and MYD88(L265P) using Gibson cloning techniques. MYD88(L265P)–13X-BioID2 was cloned by removing GFP from the previously described pBMN-MYD88(L265P)-VD-GFP[34] by restriction digest with StuI and NotI. BioID2 was PCR amplified with a 13X N-terminal linker and Gibson cloned as above from Addgene 80899 with the following primers: MYD88-Cterm/13X: CTGGACTCGC CTTGCCAAGGCCTTGTCCCTGCCCGGTGGAGGCGGGTCTGGAGGC; pBMN-NotI-BioID2-Cterm: CCTCTAGTGCGGCCGCTTATGCGTAATC CGGTACATC. BioID2 was also appended to the C terminus of both wild-type and mutant isoforms of MYD88 with a two-amino acid linker (VD). First, BioID2 and MYD88 were PCR amplified with Primestar (Takara) using the following primers: BioID FWD: TTGTCCCTGCCCGTCGACTTCAAGAAC CTGATCTGGCTG; BioID REV: CGCCGGCCCTCGAGGCTATGCGTAATC CGGTACATCG. MYD88 FWD: AATTCGAATTCCTGAAGGGCCACCATGC GACCCGACCGCGC; MYD88 REV: AGATCAGGTTCTTGAAGTCGACGGGCA GGGACAAGGC. TLR9 was cloned into a modified version of pBMN that expresses a 10× linker followed by BioID2 with the following oligonucleotides. TLR9 C-BioID FWD: CTGCCGGATCCGAATTCTAGCCACAATGGGTTTCTG CCGCAGCG; TLR9 C-BioID REV: CCCGACCCGCCTCCACCTACTTCGGCC GTGGGTCCCTGGC.

The PCR products were separated on a 1% agarose gel and column purified (Qiagen). Purified PCR products were mixed and added to pBMN-LYT2 vector that was linearized with StuI (New England Biolabs) and subjected to a Gibson reaction (New England Biolabs) following the manufacturer's protocol.

Imaging MYD88–13X-BioID2. MYD88(L265P)–13X-BioID2 was retrovirally transduced into TMD8 cells, and then purified with anti-LYT2 beads as described above. Cells were first cultured for 16 h in 50 μM biotin. Next, cells were incubated with 1 μg ml^{-1} goat anti-human IgM Fab conjugated to Alexa Fluor 488 (Jackson Immunoresearch) for 90 min at 37 °C. During this incubation period, cells were plated onto a 15 μ-Slide 8 well IbiTreat chamber slides (Ibidi) for 30 min and allowed to stick to the slides. Cells were washed twice with PBS and then fixed with 4% paraformaldehyde (Electron Microscope Sciences) for 20 min and then permeabilized with cold methanol for 10 min at −20 °C. Fixed and permeabilized cells were blocked with Duolink blocking buffer (Sigma) for 30 min at room temperature. Cells were then incubated with rabbit monoclonal antibody anti-phospho-IKKα/β (Ser176/180) (Cell Signaling Technology) diluted 1:200 in PBS with 1% BSA for 2 h at room temperature, followed by two washes with BSA/PBS. Cells were then incubated with anti-rabbit F(ab')2 conjugated to Alexa Fluor 555 (Cell Signaling Technology) at 1:1,000, mouse anti-LAMP1 conjugated to Alexa Fluor 405 (Santa Cruz Biotechnology) at 1:50 and streptavidin conjugated to Alexa Fluor 647 (Biolegend) at 1:1,000, all diluted in BSA/PBS and allowed to incubate for 1 h at room temperature. Cells were then washed for 15 min in a large volume of PBS:BSA and mounted with Prolong Diamond mounting media (Invitrogen). Images were acquired on a Zeiss LSM 880 Confocal microscope. Images for display were prepared with NIH ImageJ/FIJI[35] and animations were prepared using the Imaris 3D rendering software (Bitplane). The number of BioID2 puncta and their intensity were quantified from *z*-stack images (1 μm slices) using Blobfinder.

In certain instances, TMD8, OCI-Ly10 and/or HBL1 cells expressing MYD88(L265P)–13X-BioID2 were also transduced with either control shRNA (SC4) or shRNAs targeting *CD79A*, *TLR9* or *MYD88*, as described above. After puromycin (Invitrogen) selection, cells stained with streptavidin conjugated to Alexa Fluor 555 (Thermo Fisher Scientific) and anti-LYT2 conjugated to Alexa Fluor 647 (Biolegend). Cells were either subjected to FACS analysis, as described above, or were imaged as described above. Biotin spots or blobs were counted using Blobfinder, as for the PLA above. Likewise, these cell lines were either left untreated, treated with 10 nM ibrutinib or equivalent volumes of DMSO, and then stained and analysed in the same manner.

Mass spectrometry and western blot analysis of BioID2 constructs. TLR9–10X-BioID2 pBMN-LYT2 and MYD88–13X-BioID2 pBMN-LYT2 constructs were retrovirally transduced into DLBCL cell lines, as described above. Infected cells were enriched by positive selection with LYT2 magnetic beads (Invitrogen). Cells were then grown in SILAC media, containing arginine and lysine labelled with stable isotopes or arginine and lysine, for 2 weeks before expansion to 100×10^6 cells. In certain cases, cells were treated with either 10 nM ibrutinib or 200 nM AZD2014 for 24 h. Then, 16 h before lysis, biotin (Sigma) was added to a final concentration of 50 μM to transduced cells. Cells were then lysed at 2.5×10^7 cells per ml in RIPA buffer modified for mass spectrometry analysis (1% NP-40, 0.5% deoxycholate, 50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM Na_3VO_4, 5 mM NaF, 1 mM AEBSF) for 10 min. on ice. Lysates were cleared by centrifugation at 14,000*g* for 20 min at 4 °C. Pre-washed streptavidin agarose beads (35 μl) were added to each sample; samples were then rotated at 4 °C for 2 h, then washed four times in 1× RIPA buffer, then solubilized with 4× LDS sample buffer (Invitrogen) with 1% β-mercaptoethanol, and boiled for 5 min. A fraction of lysates were also subjected to western blot analysis as described above. Western blots were probed with rabbit anti-CARD11 and rabbit anti-MALT1 (Cell Signaling Technologies) and mouse anti-MYD88 (Santa Cruz Biotechnology).

For mass spectrometry analysis, proteins were separated by one-dimensional gel electrophoresis (4–12% NuPAGE Bis-Tris Gel) and the entire lane of a Coomassie blue-stained gel was cut into 20 slices. All slices were processed as described previously[38]. After tryptic digestion of the proteins the resulting peptides were resuspended in sample loading buffer (2% acetonitrile and 0.05% trifluoroacetic acid) and were separated by an UltiMate 3000 RSLCnano HPLC system (Thermo Fisher Scientific) coupled online to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). First, peptides were desalted on a reverse phase C18 precolumn (Dionex 5 mm length, 0.3 mm inner diameter) for 3 min. After 3 min the precolumn was switched online to the analytical column (30 cm length, 75 mm inner diameter) prepared in-house using ReproSil-Pur C18 AQ 1.9 mm reversed phase resin (Dr. Maisch GmbH). Buffer A consisted of 0.1% formic acid in H_2O, and buffer B consisted of 80% acetonitrile and 0.1% formic acid in H_2O. The peptides eluted from buffer B (5–42% gradient) at a flow rate of 300 nl min^{-1} over 76 min. The temperature of the precolumn and the analytical column was set to 50 °C during the chromatography. The mass spectrometer was operated in a TopN data-dependent mode, in which the 30 most intense precursors from survey MS1 scans were selected with an isolation window of 1.6 Th for MS2 fragmentation under a normalized collision energy of 28. Only precursor ions with a charge state between 2 and 5 were selected. MS1 scans were acquired with a mass range from 350 to 1,800 m/z at a resolution of 60,000 at 200 m/z. MS2 scans were acquired with a starting mass of 110 Th at a resolution of 15,000 at 200 m/z with maximum IT of 54 ms. AGC targets for MS1 and MS2 scans were set to 1×10^6 and 1×10^5, respectively. Dynamic exclusion was set to 20 s.

Mass spectrometry data analysis. Mass spectrometry data analysis was performed using the software MaxQuant (version 1.6.0.1) linked to the UniProtKB/Swiss-Prot human database containing 155,990 protein entries and supplemented with 245 frequently observed contaminants via the Andromeda search engine[39]. Precursor and fragment ion mass tolerances were set to 6 and 20 p.p.m. after initial recalibration, respectively. Protein biotinylation, N-terminal acetylation and methionine oxidation were allowed as variable modifications. Cysteine carbamidomethylation was defined as a fixed modification. Minimal peptide length was set to seven amino acids, with a maximum of two missed cleavages. The false discovery rate (FDR) was set to 1% on both the peptide and the protein level using a forward-and-reverse concatenated decoy database approach. For SILAC quantification, multiplicity was set to two or three for double (Lys + 0/Arg + 0, Lys + 8/Arg + 10) or triple (Lys + 0/Arg + 0, Lys + 4/Arg + 6, Lys + 8/Arg + 10) labelling, respectively. At least two ratio counts were required for peptide quantification. The 're-quantify' option of MaxQuant was enabled. Data were filtered for low confidence peptides.

Xenograft. All mouse experiments were approved by the National Cancer Institute Animal Care and Use Committee (NCI-ACUC) and were performed in accordance with NCI-ACUC guidelines and under approved protocols. Approved protocols allowed tumour growth below 20 mm in any dimension; no animals had tumours which exceeded these limits. Female NSG (non-obese diabetic (NOD)/severe combined immunodeficient (SCID)/*Il2rg*$^{-/-}$) mice were obtained from NCI Fredrick Biological Testing Branch and used for the xenograft experiments between 6 and 8 weeks of age. TMD8 tumours were established by subcutaneous injection of 10×10^6 cells in a 1:1 Matrigel/PBS suspension. Treatments were initiated when tumour volume reached a mean of 200 mm^3. Ibrutinib was prepared in PBS with 50% (v/v) DMSO and administered intraperitoneally once per day (5 mg kg^{-1} day^{-1}). AZD2014 was prepared in deionized water with 1% (v/v) Tween 80 and administered per os once per day (15 mg kg^{-1} day^{-1}). For ADZ2014 and ibrutinib combination, drugs were given at the same concentration and schedule as single agents. Tumour growth was monitored every other day by measuring tumour size in two orthogonal dimensions and tumour volume was calculated by the following equation: tumour volume = (length × width^2)/2. Treatment randomization and experimenter blinding were not possible. Sample size was estimated based on preliminary experiments. Mice were censored if they died during treatment.

FFPE biopsies. All cases were needle aspirates, whole lymph node biopsies, or were obtained from surgically removed specimens. Samples were fixed in 10% buffered formalin for 18–24 h and paraffin embedded for long term storage. Samples were studied in accordance with the ethics and principles of the Declaration of Helsinki and under Institutional Review Board approved protocols from the National Cancer Institute National Institutes of Health Protocol Review Office (protocol numbers 10-C-0181, 10-CN-074 C, 00-C-133, 00-C-133) or Johns Hopkins School of Medicine (IRB00154052). Informed consent was obtained from all patients or given an IRB-waiver as archived tissue submitted for consultation

to the Hematopathology Section. All samples were anonymized or de-identified for subsequent PLA analysis.

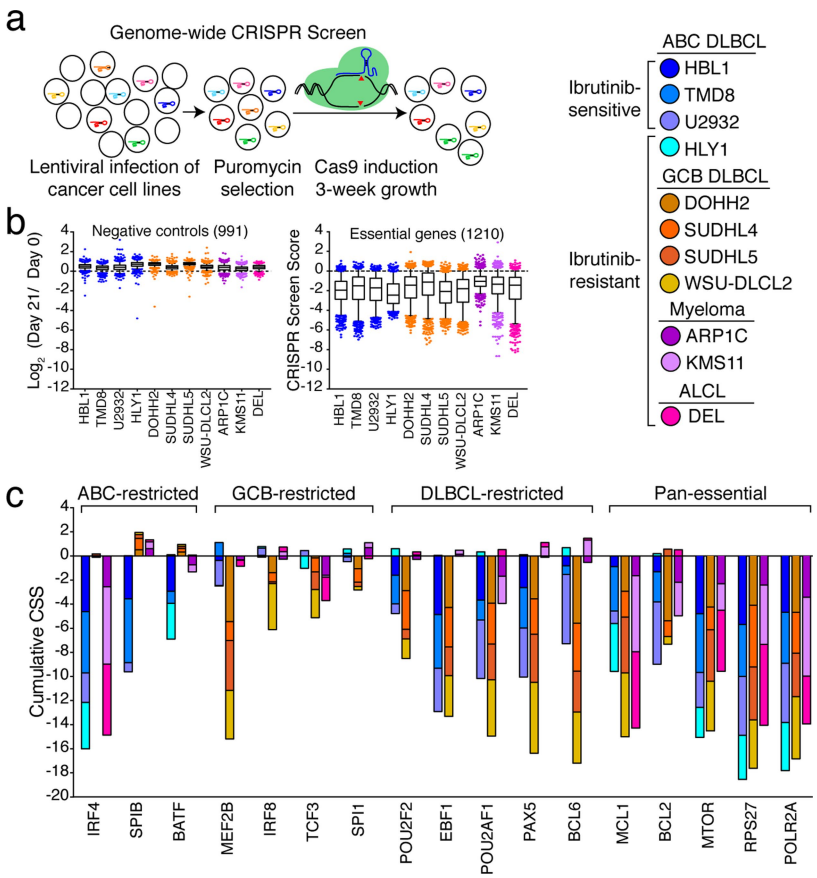
shRNA and sgRNA sequences used in functional assays. shSC4 (MSMO1 ex5) CTCTCAACCCTTTAAATCTGA; shCD19 (3′ UTR) GATTCACACCTGACT CTGAAA; shCD79A (3′ UTR) GGGGCTTCCTTAGTCATATTC; shTLR9 #1 (N133) GAGCTAAACCTGAGCTACAAC; shTLR9 #2 (3′ UTR) GCACGGTGCC ACCTCCACACT; shMYD88 #1 (3′ UTR) GTACCAGTATTTATACCTCTA; shMYD88 #2 (ex3) GGCATATGCCTGAGCGTTTC; shBCL10 #2 (3′ UTR) CTGACATTGTCTCCTATATA; shCARD11 (3′ UTR) GGGGTGTGTACCA GGCTATGA; sgTLR9 #8 GACCAGGCTCCCGAAGGAAG; sgMYD88 #10 CCGGCAACTGGAGACACAAG; sgUNC93B1 #B73 TGTTGCCATACT TCACCTCG; sgCNPY3 #9 TCAGCACGTGGTTGGCGCAG.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

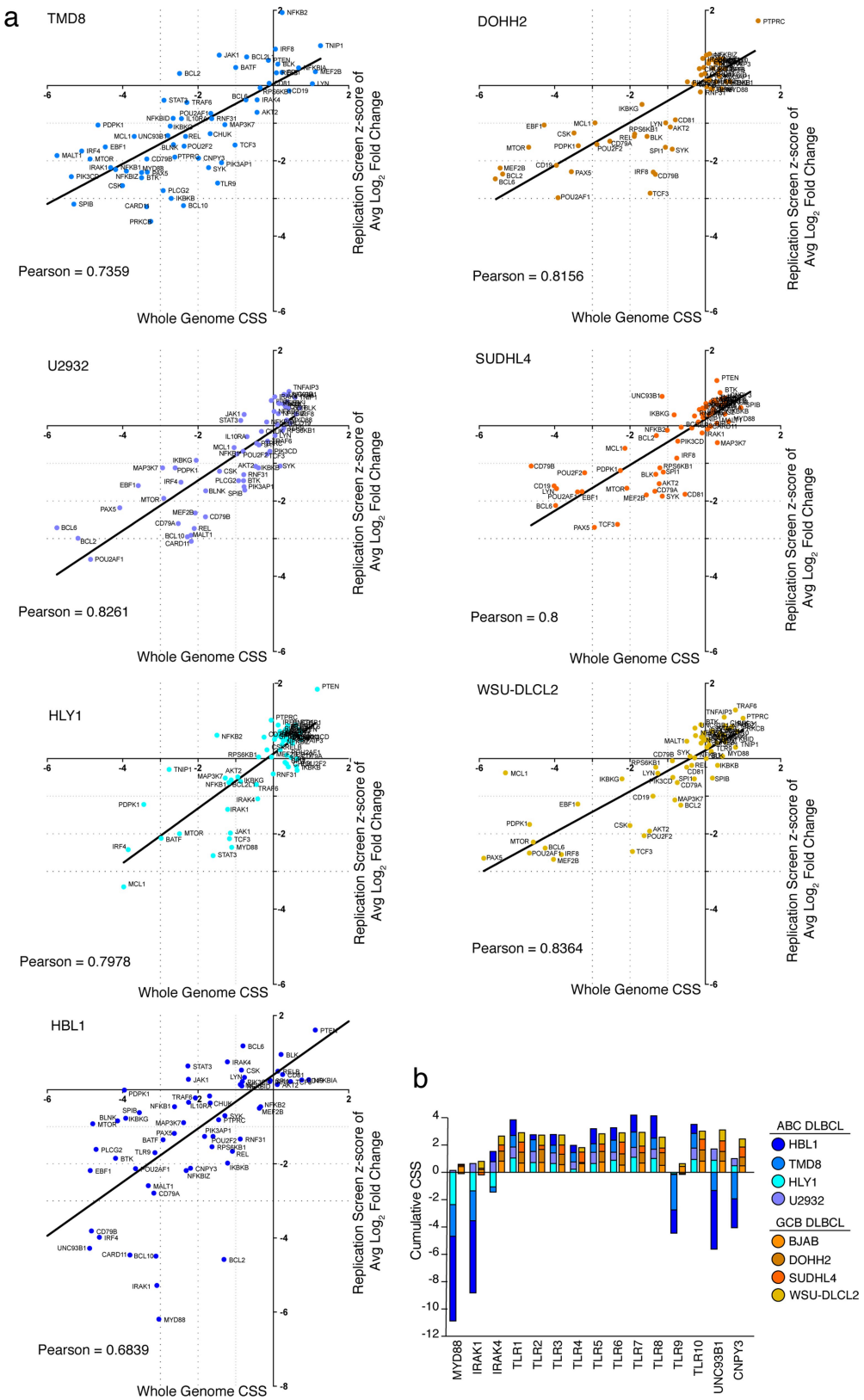
Code availability. All computer code is available at https://lymphochip.nih.gov/local/CRISPR/.

Data availability. The gene expression datasets generated for these analyses are included in the Supplementary Information, or have been deposited in Gene Expression Omnibus (GEO) under accession numbers GSE99276. Primary sequencing data and copy number analysis DLBCL cases will be made available through the NIH dbGAP system (accession numbers phs001444, phs001184 and phs000178, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001444.v1.p1) and the NCI Genomic Data Commons. All CRISPR screen data, SILAC mass spectrometry, and genomic data used in the manuscript are included in the Supplementary Tables.

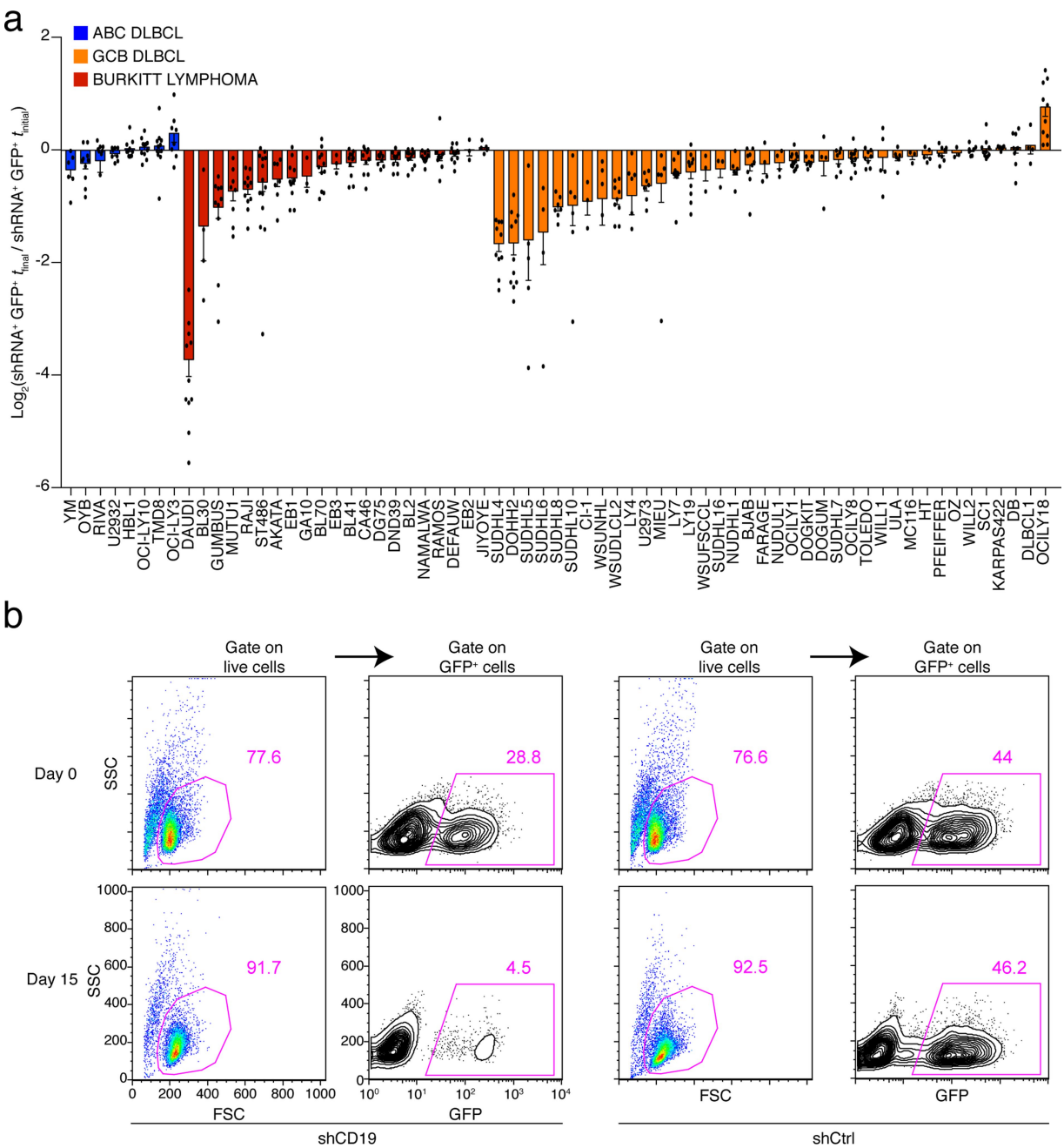
24. Brummelkamp, T. R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243–247 (2002).
25. Ngo, V. N. et al. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006).
26. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
27. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
28. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
29. Wright, G. W. & Simon, R. M. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* **19**, 2448–2455 (2003).
30. Hafner, M., Niepel, M., Chung, M. & Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* **13**, 521–527 (2016).
31. Ceribelli, M. et al. Blockade of oncogenic IκB kinase activity in diffuse large B-cell lymphoma by bromodomain and extraterminal domain protein inhibitors. *Proc. Natl Acad. Sci. USA* **111**, 11365–11370 (2014).
32. Lenz, G. et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc. Natl Acad. Sci. USA* **105**, 13520–13525 (2008).
33. Lam, L. T. et al. Small molecule inhibitors of IκB kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clin. Cancer Res.* **11**, 28–40 (2005).
34. Ngo, V. N. et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
35. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
36. Allalou, A. & Wählby, C. BlobFinder, a tool for fluorescence microscopy image cytometry. *Comput. Methods Programs Biomed.* **94**, 58–65 (2009).
37. Hewitt, S. M. The application of tissue microarrays in the validation of microarray results. *Methods Enzymol.* **410**, 400–415 (2006).
38. Oellerich, T. et al. β_2 integrin-derived signals induce cell survival and proliferation of AML blasts by activating a Syk/STAT signaling axis. *Blood* **121**, 3889–3899 (2013).
39. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).



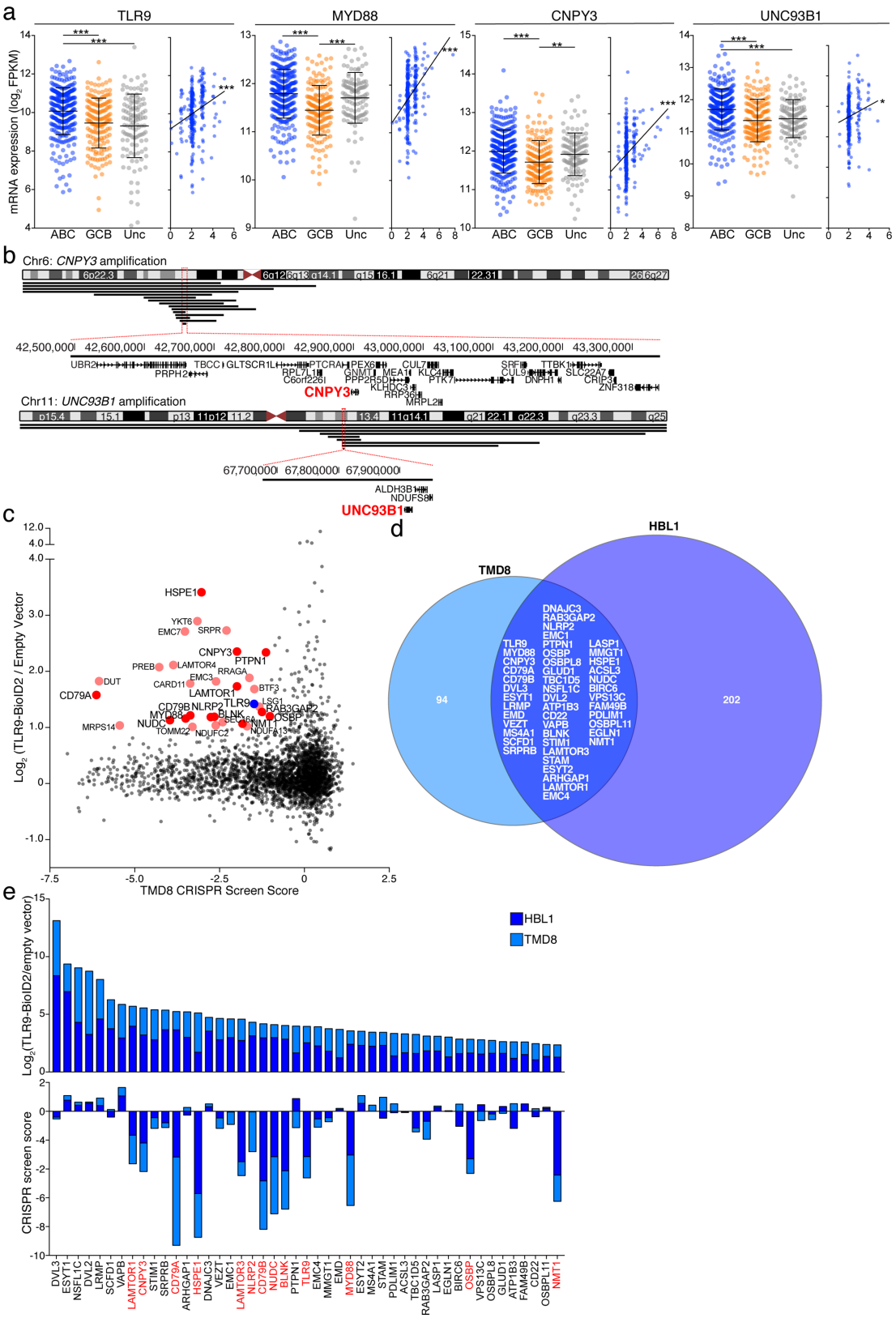
Extended Data Fig. 1 | CRISPR screen controls. a, Schematic of CRISPR–Cas9 screens in lymphoma cell lines. **b**, 991 negative control non-targeting (left) or 1,210 positive control, essential gene (right) targeting sgRNAs are displayed for each cell line with indicated metrics. Box and whisker plots display mean and interquartile data, outliers represent 10% of the total dataset. **c**, Cumulative CRISPR screen scores for indicated genes are displayed for lymphoma cell lines screened.



Extended Data Fig. 2 | Correlation of genome-wide and replication CRISPR screens. A subset of lymphoma cell lines were rescreened with replication libraries sgRNAs targeting each of the displayed 62 genes. Depletion scores of the genome-wide screen are shown on the *x* axis, and the *z*-score of the average $\log_2$ fold change of all sgRNAs targeting a given gene is shown on the *y* axis for the replication screen. Pearson correlations ($n = 62$) and linear regressions are displayed for each of the overlapping datasets. **b**, Cumulative CRISPR screen scores for TLR-pathway genes are displayed for ABC (blue) and GCB (orange) DLBCL cell lines.

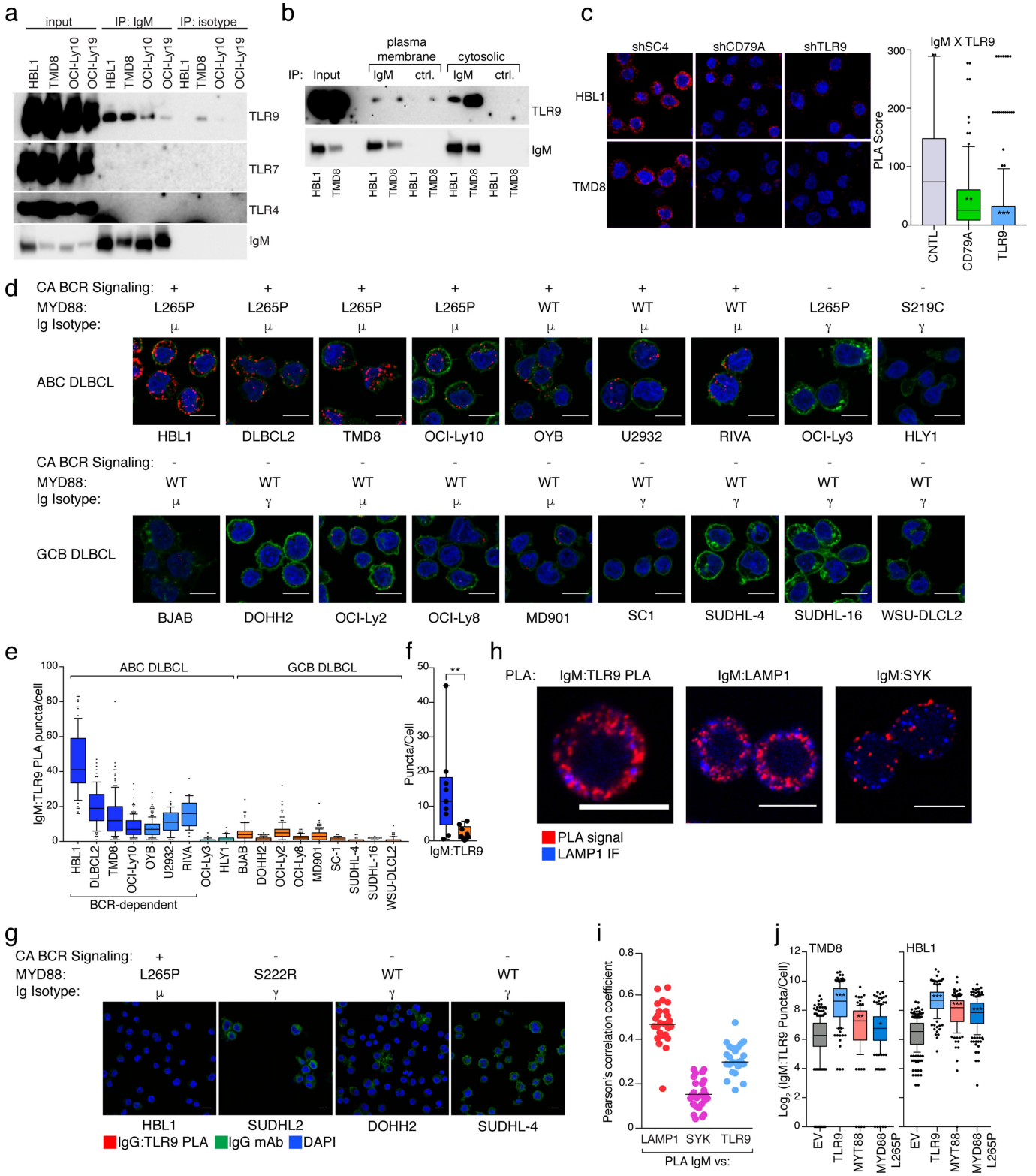


Extended Data Fig. 3 | CD19 is required for GCB and Burkitt lymphoma survival. a, A panel of 67 lymphoma cell lines was transduced with an shRNA targeting CD19. Shown is the log ratio of the percentage of live, shRNA-containing (GFP^+) cells at the last time point (t_{final}, 10–12 days) versus the intial time point ($t_{initial}$, day 0). ABC lines are depicted in blue, GCB lines depicted in orange, and Burkitt lymphoma lines are depicted in dark red. Data are mean ± s.e.m. are displayed from independent biological replicates. See 'statistics and reproducibility' section. **b**, FACS gating strategy for live, GFP^+ shRNA or sgRNA expressing cells with knockdown of CD19 or negative control genes.



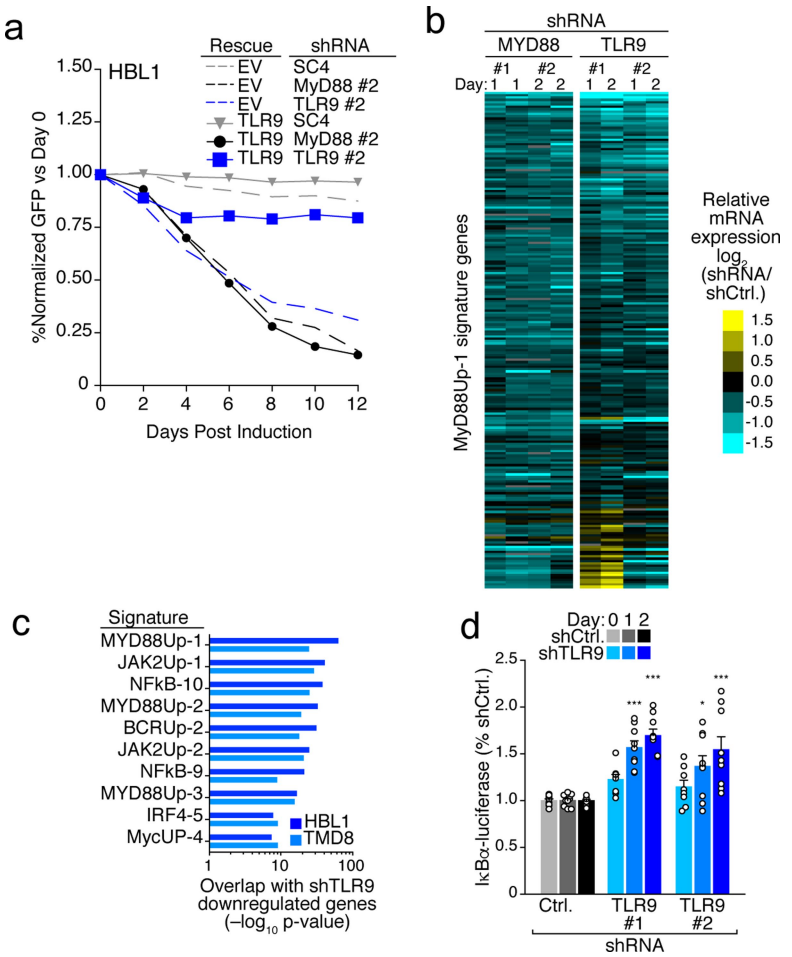
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | TLR9 overexpression and association with the BCR are features of ABC DLBCL. a, Gene expression values ($\log_2$ FPKM) values of TLR9 associated genes are shown by DLBCL subtype, ABC in blue ($n = 294$), GCB in orange ($n = 164$) and unclassified (Unc) in grey ($n = 115$). Gene expression data were correlated with DNA copy number and linear regression calculated for ABC samples. $^*P < 0.05$, $^{***}P < 0.001$, linear regression (left); $^*P < 0.05$, $^{***}P < 0.0001$, one-way ANOVA and Tukey's post test (right). **b**, Amplification of the UNC93B1 and CNPY3 loci (black lines, below chromosome ideogram). Minimal shared amplified regions in ABC DLBCL biopsies are bracketed and genes displayed below. **c**, The essential TLR9 interactome in TMD8. TLR9–BioID2 interactome determined by SILAC-based mass spectrometry (*y* axis) plotted by the CRISPR screen score (CSS, *x* axis). Bait (TLR9) is labelled in blue. Essential interactors are labelled in red, those shared with HBL1 (Fig. 2c) are labelled in dark red. **d**, Venn diagram of the overlap of TLR9–BioID2 interactors identified by SILAC-based mass spectrometry in experiments performed in TMD8 and HBL1 ABC lines. The 47 overlapping proteins are listed. **e**, The enrichment of 47 overlapping TLR9–BioID2 proximal proteins is shown (top) relative to their CSS (bottom). Gene names labelled in red are enriched and toxic to both HBL1 and TMD8. See 'statistics and reproducibility' section.

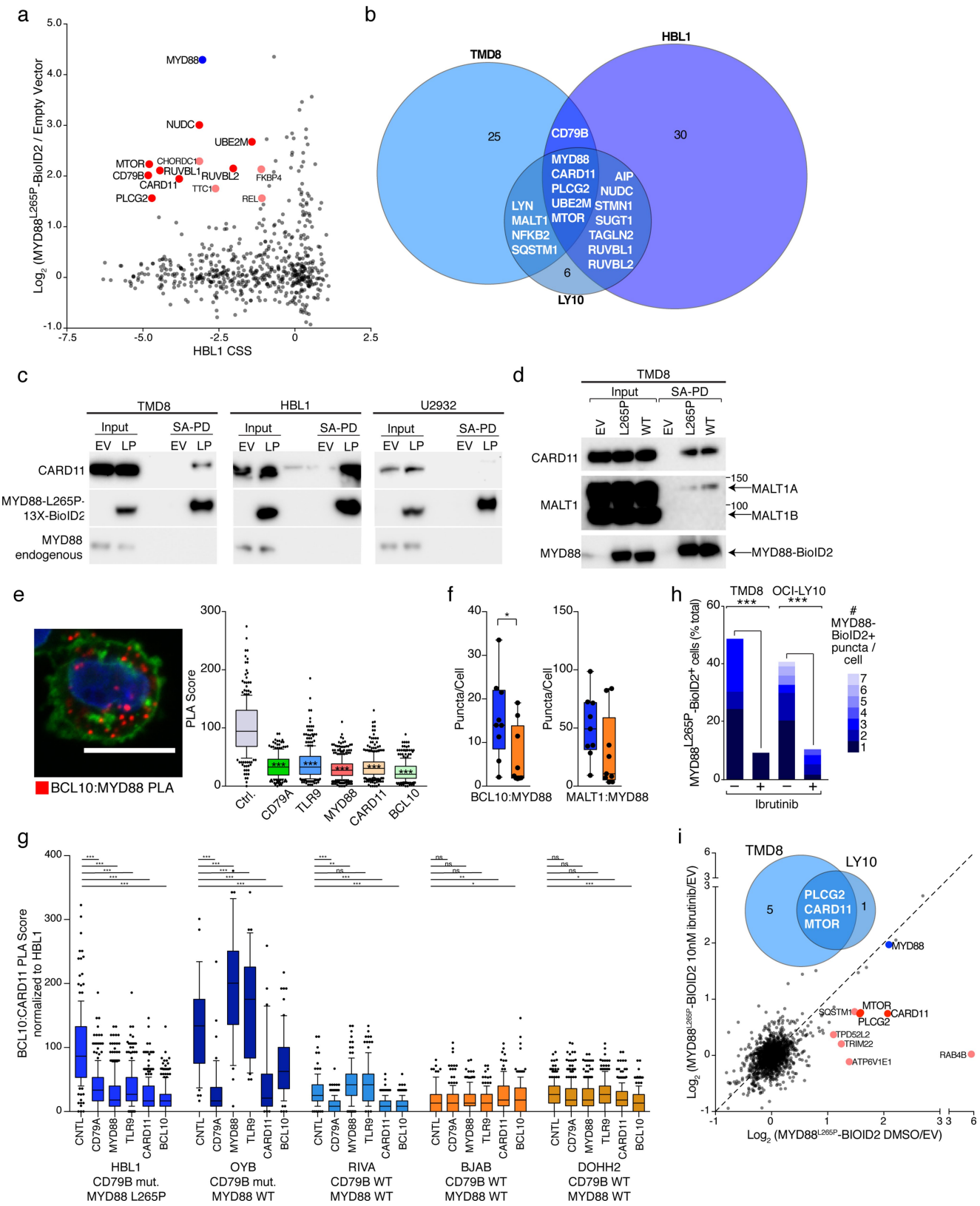


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | IgM interacts with intracellular TLR9 in ABC DLBCL lines. a, Whole-cell lysates of indicated DLBCL cell lines were immunoprecipitated with anti-IgM or isotype control antibodies before being immunoblotted with IgM or indicated TLR antibodies, representative blots; $n = 3$. **b**, ABC DLBCL cell lines HBL1 and TMD8 were incubated on ice with IgM or isotype control antibodies and lysed. Lysates were immunoprecipitated (plasma membrane) with IgM or isotype control. Unbound lysates (cytosolic) were then immunoprecipitated with IgM or isotype control antibodies before all immunoprecipitated lysates were immunoblotted with the indicated antibodies; representative blots, $n = 2$. **c**, Left, confocal images of PLA reaction between IgM and TLR9 in HBL1 and TMD8 cells transduced with control SC4, *CD79A* or *TLR9* shRNAs. Cells were puromycin selected and shRNAs induced with dox for two days before being fixed and imaged. Right, quantification data from three separate experiments. Data are pooled biologically independent experiments of $n > 100$ cells scored per condition. Box plots represents median and 25–75% of data, whiskers display 10–90 percentile. $^{**}P < 0.001$, $^{***}P < 0.001$, one-way ANOVA with Dunnett's post test. **d**, An IgM–TLR9 PLA (red) was performed in a panel of ABC and GCB DLBCL cell lines and the presence of chronic active BCR signalling ('+' denotes present), MYD88 mutational status and IgH isotype (μ = IgM, γ = IgG) are displayed. Nuclei were stained with DAPI (blue) and membranes were visualized by WGA (green). **e**, The number of puncta per cell of IgM–TLR9 PLA is quantitated. Box and whisker plots display mean and interquartile data, whiskers display 10–90 percentile. Data are from three fields of cells quantified from one representative experiment of three biologically independent replicates. **f**, The data from Extended Data Fig. 5e segregated by ABC (blue, $n = 9$) and GCB (orange, $n = 9$) lines. Box plots represent median and 25–75% of data, whiskers display range. $^{**}P < 0.01$, Mann–Whitney unpaired one-tailed *t*-test. **g**, IgG–TLR9 PLA (red) was performed in indicated DLBCL cell lines co-stained with DAPI (blue) and IgG-AlexaFluor488 (green). MYD88 mutational status, IgH isotype and presence of chronic active BCR signalling ('+' denotes present; '–' denotes absent) are displayed. Representative data from two independent biological replicates. **h**, To define the cytoplasmic location of the BCR–TLR9 interaction, we counterstained ABC cells for LAMP1, a marker of late endolysosomes, in which TLR9 resides, and performed PLA between IgM–TLR9, IgM–LAMP1 and IgM–SYK. The PLA signal is in red, LAMP1 is counterstained in blue. Representative data from three independent biological experiments. **i**, To quantify the association between PLA signals and LAMP1 staining, we calculated the Pearson's correlation coefficients across all pixels in each imaged cell ($n = 25$ cells per PLA pair). The highest correlation was between an IgM–LAMP1 PLA and LAMP1 staining ($R = 0.471$), whereas the correlation between an IgM–SYK PLA signal and LAMP1 was much lower ($R = 0.153$). The correlation between the IgM–TLR9 PLA signal and LAMP1 staining was intermediate ($R = 0.310$), indicating that a significant component of the IgM–TLR9 interaction is in $LAMP1^+$ vesicles. Quantified data are from one of three independent biological experiments. **j**, Quantification of the IgM–TLR9 PLA signal after ectopic expression of either empty vector, TLR9, wild-type MYD88 or MYD88(L265P). Data pooled from 3 (HBL1) or 2 (TMD8) biologically independent replicates of $n \geq 100$ cells scored per condition. Box plots represents median and 25–75% of data, whiskers display 10–90 percentile. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; one-way ANOVA with Dunnett's post test. See 'statistics and reproducibility'.

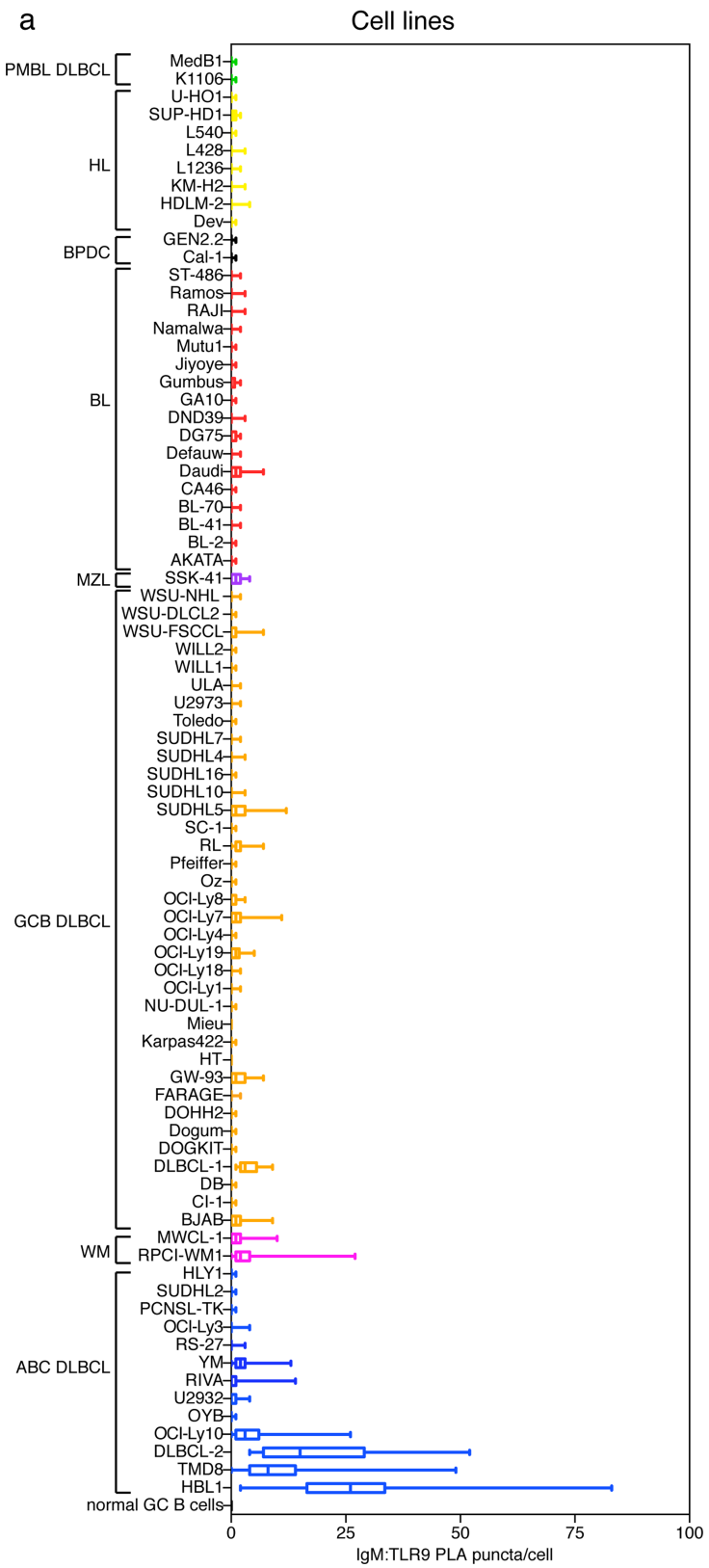


Extended Data Fig. 6 | TLR9 knockdown phenocopies MYD88 knockdown. a, TLR9 shRNA is rescued by overexpression of TLR9. HBL1 cells were transduced with empty vector (EV) or wild-type TLR9 expressing dsRedExpress2 vectors and then with shRNA vectors marked by GFP targeting a control (SC4), *MYD88* or *TLR9*. The percentage of double-positive cells was monitored by FACS and normalized to day 0. One of three representative biologically independent experiments is shown. **b**, Heat map of gene expression values showing the global phenocopy of MYD88-dependent genes after shRNA-mediated knockdown of TLR9 or MYD88 in HBL1 at indicated time points. **c**, Gene signatures enriched in downregulated genes from HBL1 or TMD8 after shRNA-mediated knockdown of TLR9. **d**, Normalized IκBα−luciferase reporter levels at indicated time points after knockdown of TLR9 with indicated shRNAs. Data are mean and s.e.m. of nine technical replicates from $n = 3$ independent biological experiments., $^{*}P < 0.05$, $^{***}P < 0.001$, one-way ANOVA with Sidak's multiple comparison test.

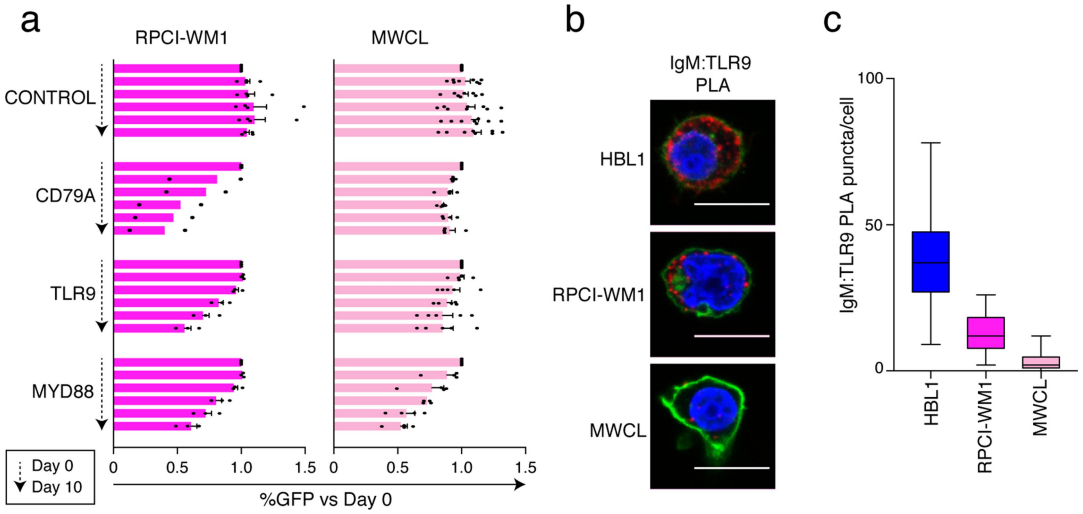


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | The MYD88(L265P) interactome in ABC DLBCL. a, The essential MYD88(L265P) interactome in HBL1. MYD88(L265P) –BioID2 interactome from SILAC-based mass spectrometry (*y* axis) plotted by the CSS (*x* axis). Bait (MYD88(L265P)) is labelled in blue. Essential interactors are red, with those shared with either TMD8 or OCI-Ly10 labelled in dark red. **b**, Venn diagram of the overlap of MYD88(L265P)–BioID2 interactors in TMD8, OCI-Ly10 and HBL1 ABC lines. Proteins found in two or more experiments are listed. **c**, Lysates of TMD8, HBL1 and U2932 cells transduced with empty vector or MYD88(L265P)–BioID2, selected and treated with 50 μM biotin for 24 h. Lysates were prepared and immunoprecipitated with streptavidin before being immunoblotted with CARD11 and MYD88 antibodies. One representative blot is shown for each cell line from $n = 3$ biologically independent experiments (HBL1, TMD8) and $n = 1$ (U2932) **d**, Lysates of TMD8 cells transduced with empty vector, MYD88(L265P) or wild-type BioID2-fusion proteins, selected and treated with 50 μM biotin for 24 h. Lysates were prepared and immunoprecipitated with streptavidin before being immunoblotted with CARD11, MALT1 or MYD88 antibodies; representative blot; $n = 3$. **e**, Confocal image of a PLA of BCL10 with MYD88. Data pooled from 6 biologically independent replicates of $n > 200$ cells scored per condition. Box plots represent median and 25–75% of data, whiskers display 10–90 percentile; one-way ANOVA with Dunnett's post test. **f**, BCL10–MYD88 and MALT1–MYD88 PLA in ABC (blue, $n = 9$) and GCB (orange, $n = 9$) lines. Box plots represent median and 25–75% of data, whiskers display range; Mann–Whitney unpaired, one-tailed *t*-test. **g**, BCL10–CARD11 PLA after shRNA knockdown of indicated genes in ABC (blue) and GCB (orange) lines. CD79B and MYD88 mutation status is displayed below each cell line. Data are from 3 fields of cells quantitated from 1 representative experiment of 3 (HBL1), 2 (BJAB, DOHH2) or 1 (OYB, RIVA) biologically independent replicates of $n \geq 90$ cells scored per condition. Box plots represent median and 25–75% of data, whiskers display 10–90 percentile; one-way ANOVA with Dunnett's post test. **h**, ABC lines expressing MYD88(L265P)–BioID2 were treated with DMSO or 10 nM ibrutinib for 24 h, and the numbers of biotin puncta were quantified from confocal images (representative experiment, $n = 3$). Fisher's exact test, two-sided. **i**, SILAC-based mass spectrometry comparison of MYD88(L265P)–BioID2 interactome in TMD8 cells treated with DMSO (*x* axis) versus 10 nM ibrutinib (*y* axis). Proteins reduced upon ibrutinib treatment are shown in red, those similarly decreased in two separate cell lines (Fig. 4a) are labelled in dark red. Bait (MYD88) is labelled in blue. Venn diagram showing overlap of proteins decreased by more than 30% in OCI-Ly10 cells (Fig. 4a) is shown as an inset. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; see 'statistics and reproducibility' section.



Extended Data Fig. 8 | IgM–TLR9 PLA identifies ABC samples with chronic active BCR signalling in tissue microarrays. a, IgM–TLR9 PLA was performed on a formalin-fixed, paraffin-embedded (FFPE) tissue microarray of lymphoma cell lines. PLA puncta were quantified and plotted as the absolute number of spots per cell from at least 2 images of 1 representative experiment from 3 independent tissue microarray replicates. Box plots represent median and 25–75% of data, whiskers display range. Cell lines are divided by putative lymphoma subtype for presentation. BL, Burkitt lymphoma; BPDC, blastic plasmacytoid dendritic cell neoplasm; HL, Hodgkin lymphoma; MZL, marginal zone lymphoma; PMBL, primary mediastinal B cell lymphoma; WM, Waldenström's macroglobulinemia. **b**, Representative confocal fluorescent image from three independent biological samples of a germinal centre from a reactive lymph node. IgM–TLR9 PLA is shown in red; CD20 is in green; CD138 is in white; and DAPI is in blue.



Extended Data Fig. 9 | Waldenström's macroglobulinaemia can utilize the My-T-BCR supercomplex. a, shRNA-mediated toxicity of indicated genes in two Waldenström's macroglobulinaemia cell lines (RPCI-WM1 and MWCL-1). Control (SC4), *CD79A*, *TLR9* or *MYD88* shRNAs were expressed in tandem with GFP and the relative level of GFP was followed over time by FACS. Data are mean and s.e.m. of independent biological experiments; see 'statistics and reproducibility' section. **b**, Confocal images from one of two representative biologically independent experiments of the PLA reaction between IgM and TLR9 (red puncta). Cells were counterstained with DAPI (blue) and WGA (green). Scale bars, 10 μm. **c**, Normalized quantification (PLA score) of IgM–TLR9. Data were quantified from at least 28 cells per condition. Box plots represent median and 25–75% of data, whiskers display range.

LETTER

https://doi.org/10.1038/s41586-018-0374-x

Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast

Jingchuan Luo[1,2], Xiaoji Sun[1], Brendan P. Cormack[2] & Jef D. Boeke[1]*

Extant species have wildly different numbers of chromosomes, even among taxa with relatively similar genome sizes (for example, insects)[1,2]. This is likely to reflect accidents of genome history, such as telomere–telomere fusions and genome duplication events[3–5]. Humans have 23 pairs of chromosomes, whereas other apes have 24. One human chromosome is a fusion product of the ancestral state[6]. This raises the question: how well can species tolerate a change in chromosome numbers without substantial changes to genome content? Many tools are used in chromosome engineering in *Saccharomyces cerevisiae*[7–10], but CRISPR–Cas9-mediated genome editing facilitates the most aggressive engineering strategies. Here we successfully fused yeast chromosomes using CRISPR–Cas9, generating a near-isogenic series of strains with progressively fewer chromosomes ranging from sixteen to two. A strain carrying only two chromosomes of about six megabases each exhibited modest transcriptomic changes and grew without major defects. When we crossed a sixteen-chromosome strain with strains with fewer chromosomes, we noted two trends. As the number of chromosomes dropped below sixteen, spore viability decreased markedly, reaching less than 10% for twelve chromosomes. As the number of chromosomes decreased further, yeast sporulation was arrested: a cross between a sixteen-chromosome strain and an eight-chromosome strain showed greatly reduced full tetrad formation and less than 1% sporulation, from which no viable spores could be recovered. However, homotypic crosses between pairs of strains with eight, four or two chromosomes produced excellent sporulation and spore viability. These results indicate that eight chromosome–chromosome fusion events suffice to isolate strains reproductively. Overall, budding yeast tolerates a reduction in chromosome number unexpectedly well, providing a striking example of the robustness of genomes to change.

Chromosome engineering in *S. cerevisiae* is driven by technological advances[7–10]. A haploid yeast strain with 33 chromosomes has been generated by splitting natural chromosomes into smaller chromosomes[7]. On the other hand, the two largest *S. cerevisiae* chromosomes, IV and XII, were fused by homologous recombination[8], producing yeast with a 3.2-Mb compound chromosome that grew well. Further fusions (chromosomes VII–V–XV–IV) generated a 4.3-Mb compound chromosome with the longest yeast chromosome arm engineered previously (3.7 Mb), and a haploid chromosome number of $n = 12$, with no observed effect on fitness[9]. CRISPR–Cas9 expression in *S. cerevisiae*[10] permits efficient engineering without selection. We used this to push the lower limit of chromosome numbers in *S. cerevisiae* by fusing chromosomes, producing a series of strains with progressively fewer chromosomes without affecting gene content.

At least three potential biological obstacles might hinder engineering of karyotype. First, studies of the field bean *Vicia faba* have suggested that the length limit to chromosome arms is half the average spindle axis at telophase for normal development[11]. Longer arms might yield incomplete sister chromatid separation, lagging chromosomes and micronucleus formation, impairing fertility and development. Second, as centromeres are deleted, excess kinetochore proteins may cause problems. Finally, mitotic mechanics may be affected—centromeric force may not suffice to pull very long chromosomes poleward. The latter is a particular concern for *S. cerevisiae*, in which point centromeres are bound by single microtubules; most organisms with larger chromosomes have regional centromeres bound by multiple microtubules[12]. Karyotype engineering in *S. cerevisiae*, therefore, investigates whether point centromeres can segregate large chromosomes.

We devised specific paths to evaluate minimization of the number of chromosomes (n). We first fused all small chromosomes to maximize the number of chromosomes fused before hitting a potential chromosome arm length limit. Specifically, we fused chromosomes IX, III and I, then V and VIII, and finally II and VI to generate an $n = 12$ strain (Fig. 1).

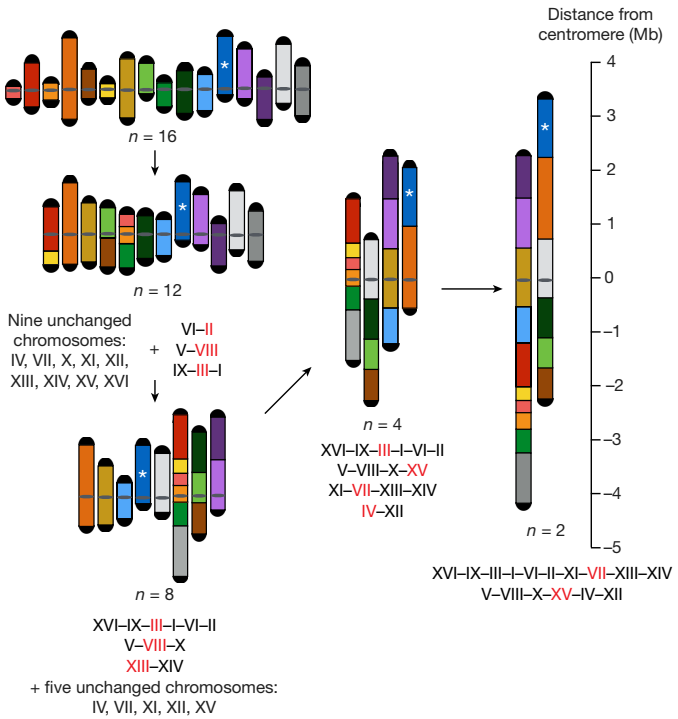


Fig. 1 | Fusion chromosome paths and strategy. This diagram shows how we fused chromosomes together from the wild type ($n = 16$) to $n = 12$, $n = 8$, $n = 4$ and finally $n = 2$. The 16 chromosomes are coloured uniquely and arranged by number. A ruler indicates the distance from the centromere. The centromeres of $n = 4$ and $n = 2$ are aligned to the 0 position. Please note that the length of the rDNA array, whose position is indicated with an asterisk, is omitted. Red lettering indicates the chromosome that has an active centromere in the compound chromosome. All the chromosomes are oriented from left to right (bottom to top in the figure) in the fusion chromosomes. For more details see Extended Data Fig. 1b.

[1]Institute for Systems Genetics, NYU Langone Health, New York, NY, USA. [2]Department of Molecular Biology & Genetics, JHU School of Medicine, Baltimore, MD, USA. *e-mail: jef.boeke@nyumc.org

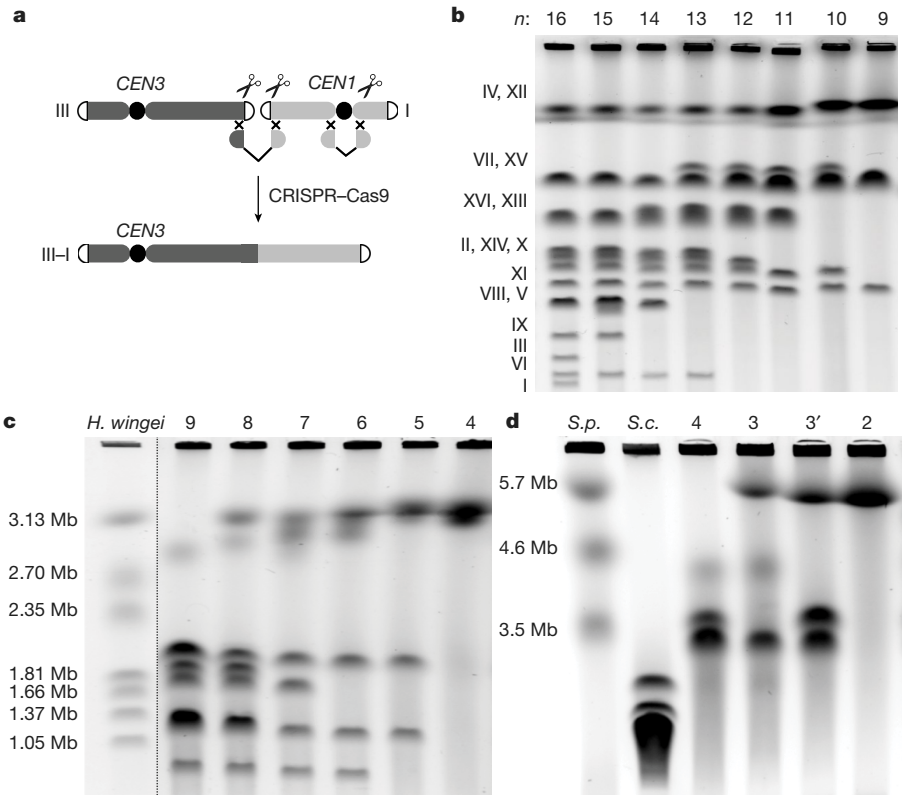


Fig. 2 | Construction and characterization of fusion chromosomes. **a**, A schematic showing a CRISPR–Cas9 based method to fuse any two chromosomes together. CEN3 and CEN1, centromeres of chromosomes III and I, respectively. **b**, Pulsed-field gel electrophoresis with a standard protocol for *Saccharomyces cerevisiae*. **c**, Pulsed-field gel electrophoresis with *Hansenula wingei* (also known as *Wickerhamomyces canadensis*) chromosomal DNA as a marker. **d**, Pulsed-field gel electrophoresis with *S. pombe* (*S.p.*) chromosomal DNA as a marker. *S.c.*, *S. cerevisiae*. $n = 3$ strain is yJL381; $n = 3'$ strain is yJL410.

We developed a CRISPR–Cas9-based strategy to fuse any two chromosomes while largely preserving isogenicity (Fig. 2a, Extended Data Fig. 1a and Extended Data Table 1; see Methods). Pulsed field gel electrophoresis, used for karyotyping *S. cerevisiae*, confirmed the presence of progressively fewer and larger chromosomes (Fig. 2b) with all chromosomes shorter than 600 kb merged in the $n = 12$ strain.

Given that 3.2-Mb and 4.3-Mb compound acrocentric chromosomes have been engineered previously[8,9], we next produced a strain with four long chromosomes, each about 3 Mb long (for simplicity, the length of the rDNA array is omitted from lengths reported herein; Fig. 1). The serial disappearance of chromosomes confirmed that sequential fusion events occurred as planned (Fig. 2c). The $n = 4$ strain had four chromosomes, each around 3 Mb (Fig. 2d). At $n = 4$, the total chromosome number drops substantially below the smallest number previously achieved in engineered *S. cerevisiae* strains. Moreover, $n = 4$ is also lower than is seen in any extant *Saccharomycetaceae* species with point centromeres[3].

Finally, to generate a strain with only two metacentric chromosomes, each about 6 Mb long, we planned to fuse the final four chromosomes in any viable order or orientation (Extended Data Fig. 1c), generating two versions of $n = 3$ strains by fusing different chromosomes in the $n = 4$ strain (Fig. 2d). To distinguish between these, we refer to them as $n = 3$ (yJL381) and $n = 3'$ (yJL410) (Extended Data Table 2). A further fusion cycle in $n = 3$ generated an $n = 2$ strain, with two metacentric chromosomes each about 6 Mb long (Fig. 2d). Each chromosome in the $n = 2$ strain now carries about half of the genomic content, which is unchanged compared to the $n = 16$ strain, except for the deletion of 14 centromeres and 28 telomeres.

We tried to generate an $n = 1$ strain by multiple strategies, for example, changing arm length (balanced versus dissimilar), position of centromere, and remaining telomeres (Extended Data Fig. 1d), but were unable to produce one. Has the limit of chromosome arm length been reached? Given that anaphase spindle axis length could extend up to 10 μm[13], along with extrapolation from previous measurements (0.85 μm per 1 Mb)[8], we estimate maximum arm length at about 5.9 Mb (see Methods), or slightly less than half of the full genome length if rDNA is included. This may make it more difficult to produce an $n = 1$ strain compared to other fusion steps, because the arm length of a metacentric single chromosome may approach the maximum. However, as we sampled only a few chromosome fusion paths (of 10^{19} possible), other paths might lead to an $n = 1$ strain. An accompanying study succeeded in producing an $n = 1$ strain using related methods, in combination with deleting repetitive regions[14].

We analysed growth and resistance to various stresses for different n strains. Unexpectedly, strains with between four and sixteen chromosomes grew well in different media and stress conditions (Extended Data Fig. 2). Like $n = 4$, the $n = 3$ and $n = 2$ strains lacked obvious fitness defects (Fig. 3a). These strains are healthy, indicating that *S. cerevisiae* can handle large chromosomes and therefore, a regional centromere is not required to segregate *S. pombe*-sized chromosomes. To quantify small differences in fitness, we performed quantitative competitive growth assays by co-culturing differentially tagged BY4741 cells and $n = 4$ or $n = 2$ cells, and measuring changes in fluorescence ratios with time. The $n = 4$ strain grew similarly to the wild type (98.7 ± 0.3%), whereas the $n = 2$ strain grew slightly more slowly (91.3 ± 0.4%; Fig. 3b).

Fusion of chromosomes might trigger secondary genome or transcriptome changes. Sequencing of genomes from the $n = 2$, 4, 8 and 12 strains revealed no evidence of aneuploidy or regional copy number differences, and few new single nucleotide polymorphisms (SNPs) or small insertions/deletions (indels) (Extended Data Fig. 3 and Extended Data Table 3). The numbers of SNPs detected were within threefold of expected numbers based on reported error rates[15] (Extended Data Fig. 3c). Two out of eleven SNPs (*AMN1* L317F and *QDR3* P586S) were predicted to be deleterious by the Sorting Intolerant From

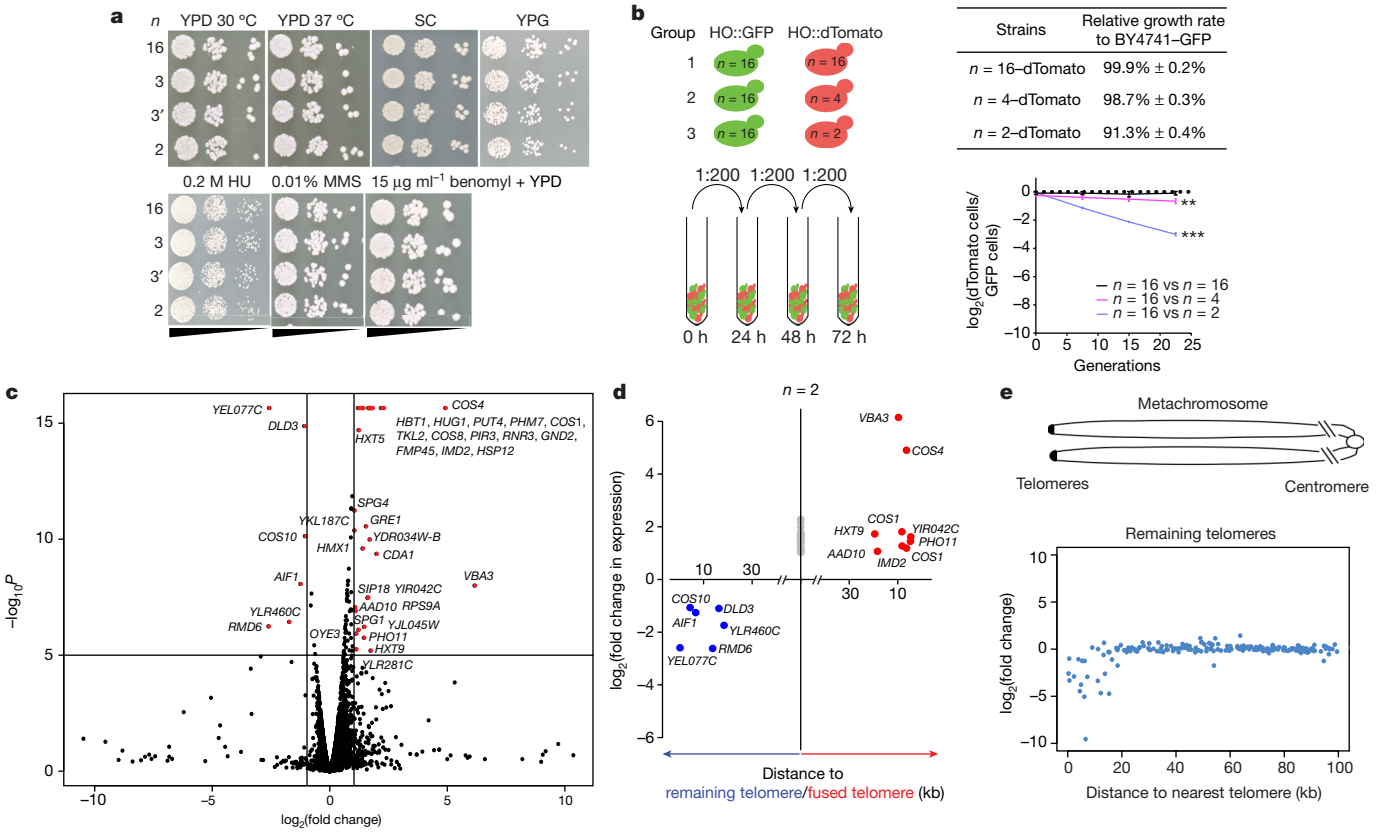


Fig. 3 | Characterization of fusion chromosome strains. a, Serial dilution assays under seven different conditions with $n=16$, $n=3$ and $n=2$ strains. HU, hydroxyurea; MMS, methyl methanesulfonate; SC, synthetic complete medium; YPG, yeast extract peptone with 3% glycerol. **b**, Competitive growth assays. Each experimental group was tested in biological triplicate. Group 2 ($n=16$ vs $n=4$), $P=0.0048$; group 3 ($n=16$ vs $n=2$), $P=0.00013$ (one-sided t-test). *$0.01<P<0.05$, **$0.001<P<0.01$, ***$P<0.001$. **c**, A volcano plot showing RNA-seq data by comparing the transcriptomes of the $n=2$ and $n=16$ strains. Red dots indicate genes whose expression was significantly different in the $n=2$ strain compared to the $n=16$ strain ($P<10^{-5}$, |fold change| > 2). Some DNA replication stress response genes, including *RNR3* and *HUG1*, are also upregulated in the $n=2$ strain. P values derived from two-sided t-test. *YKL187C* also known as *FAT3*; *YLR281C* also known as *RSO55*. **d**, Comparison of transcriptomes of $n=2$ and $n=16$ strains. Genes located within 20 kb of remaining telomeres are shown in blue; those within 20 kb of fused telomeres are shown in red; others are shown in grey. **e**, 'Metachromosome' of remaining telomere plots. Top, schematic diagram showing in the metachromosome view, all telomeres are aligned on the left; we show expression changes of genes within 100 kb of the nearest telomere. y-axis, $\log_2$(fold change in gene expression comparing transcriptome of $n=2$ to $n=16$); x-axis, distance of genes from the closest telomere.

Tolerant (SIFT) online tool (http://sift.jcvi.org/), which assumes that important positions in coding sequences of proteins are conserved during evolution. *AMN1* L317F is a potential suppressor of fusion chromosome state, because *AMN1* controls mitotic exit[16]. Fusion efficiency did not drop markedly as n was reduced (Extended Data Table 1). RNA sequencing (RNA-seq) analysis revealed only minimal perturbations in the $n=8$ and $n=4$ strains (Extended Data Table 4). Because $n=2$ does have a growth defect, we were unsurprised to find a few substantial transcriptome changes, including six downregulated genes (Fig. 3c and Extended Data Table 4). Strikingly, all these genes are positioned near the four remaining telomeres (Fig. 3d). As expression of the silencing genes *SIR2, SIR3* and *SIR4* was similar between $n=2$ and wild-type strains, we hypothesize that downregulation of these six genes was due to enhanced telomere position effect (TPE). Consistent with this, further analysis of gene expression near the remaining telomeres showed reduced expression for genes within 20 kb of the telomeric repeats (Fig. 3e). The $n=2$ strain also showed some upregulated genes lying close to newly fused telomeres; after fusion, when these genes are no longer subject to TPE, they are likely to be transcriptionally de-repressed (for example, genes in chromosomes IXR, VIL, VIIL and VIIIR; Extended Data Fig. 4). For the $n=4$ and $n=8$ strains, the few genes with substantially altered expression were also telomere-proximal (Extended Data Table 4). The significantly expressed subtelomeric genes overlapped extensively with subtelomeric genes upregulated in *sir* deletion strains (Extended Data Table 4). Finally, subtelomeric genes were highly enriched in genes whose expression was significantly affected in the $n=2$, $n=4$ and $n=8$ strains (Extended Data Table 4).

Karyotype engineering to reduce chromosome number without affecting gene content revealed only mild changes in gross phenotypic and global gene expression. It is well known that karyotypic changes, including chromosomal rearrangement, polyploidy and hybridization, contribute to post-zygotic isolation, especially in plants[17]. About 100 million years ago, a whole-genome duplication (WGD) event[4,5] resulted in a modal yeast chromosome number shift from $n=8$ to $n=16$. The major mechanism of chromosome number reduction in post-WGD species with $n<16$ and non-WGD species with $n<8$ was inferred to be via telomere fusion and loss of one centromere[3], resembling the process intentionally engineered here.

Translocation events have been associated with reproductive isolation in *Saccharomyces* species[18–21]. A single naturally occurring reciprocal translocation was estimated to reduce spore viability to 50% when backcrossed to the wild type, owing to unbalanced segregation of essential genes on those translocation regions, while non-reciprocal translocation was 75%[20]. For chromosome–chromosome fusion, it is more difficult to model the effects on spore viability. The question remains, how much variation in chromosome number results in reproductive isolation (operationally defined here as <1% viability). Here, we are

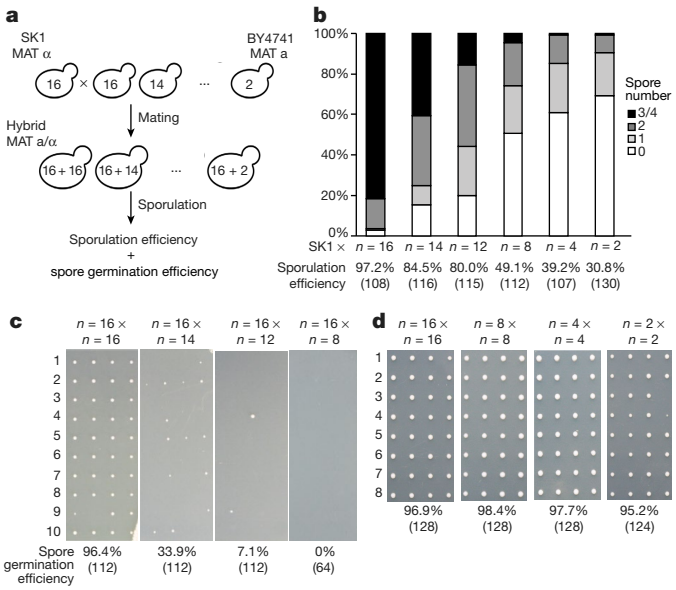


Fig. 4 | Karyotype engineering leads to reproductive isolation. **a**, Schematic of the experimental approach. The number inside the yeast is the number of chromosomes. **b**, Sporulation efficiency of hybrid strains. *Y*-axis, percentage of asci with 0–4 spores; *x*-axis, diploid strains. The number of asci counted is shown in parentheses. **c**, Spore clone germination rates for hybrid strain. Representative images are shown. The number of spores counted is shown in parentheses. **d**, Spore viability for wild-type diploid BY4743 $2n=32$, isogenic diploids: $2n=16$ strain, $2n=8$ strain and $2n=4$ strain. Representative images are shown. The number of spores counted is shown in parentheses.

able to empirically assess the impact of chromosome number in isolation, because the strains represent an isogenic series (see Methods for definition of isogenic as used here) with the parent strain, BY4741.

To explore reproductive isolation in this context, we mated each BY4741-derived strain from the fusion chromosome series ($n=16$ to $n=2$) to SK1 ($n=16$), an efficient sporulating strain, and sporulated the diploids (Fig. 4a). All crosses readily produced viable zygotes, showing no barrier to zygote formation. We confirmed that nuclear fusion had occurred (Extended Data Fig. 5a), ruling out a block in karyogamy. We examined both success rate for meiosis, that is, sporulation efficiency, and the successful segregation of one genome's worth of information into each spore, monitored by germination efficiency. For $n=16\times n=16$ crosses, more than 80% of asci developed 3–4 spores, and overall spore efficiency (per cent of asci with at least one spore) was 97.2%. For the 'asymmetric' ($n<16\times n=16$) crosses, as n decreased, sporulation efficiency dropped markedly. For $n=8\times n=16$ matings, fewer than 5% of asci developed 3 or 4 spores. For $n=2$ or $n=4\times n=16$ matings, fewer than 1% of asci produced 3 or 4 spores (Fig. 4b).

To test germination of offspring spore clones, we used a fully isogenic configuration, because spore clones that arise from $n=16\times n=16$ BY × SK1 hybrid diploids showed differences in spore clone colony size, reflecting differences between the BY and SK1 backgrounds. We crossed BY4741 strains ($n=16$–8) to BY4742 (the isogenic $n=16$ *MAT*α partner of BY4741), and performed dissections of four-spored tetrads. Whereas $n=16\times n=16$ gave rise to more than 90% viable spore clones, $n=16\times n=14$ produced only 33.9% viable spores. In $n=16\times n=12$, survival dropped markedly to below 10% (Fig. 4c, Extended Data Fig. 5c). In $n=16\times n=8$, only very few tetrads formed after 15 days' sporulation. We dissected 16 such tetrads and no spores (of 64) were viable. This did not reflect defects in chromosome fusion strains per se, as isogenic $n=8\times n=8$ diploids traverse meiosis as well as $n=16$ diploids (Extended Data Fig. 5b), giving rise to 98.4% viable spores (Fig. 4d). Indeed, isogenic $n=4\times n=4$ and $n=2\times n=2$ diploids also sporulated well and generated viable spores (Fig. 4d, Extended Data Fig. 5b). We observed 5.6% small colonies in spores from the $n=2\times n=2$ cross only. Thus, the $n=8$, 4 and 2 strains are fully capable of performing meiosis and generating viable progeny, but are incapable of generating viable progeny with the $n=16$ strain. These results indicate that eight chromosome–chromosome fusion events suffice to result in virtually complete reproductive isolation. Reproductive isolation comes from at least three sources. 1) Owing to problems with nondisjunction in meiosis I, there is a steady rise in probability that a spore will inherit a genome missing at least one chromosome's worth of genetic information as n drops. 2) Even if a full complement of genes is inherited, the probability of a lethal dosage imbalance is elevated in asymmetric crosses. 3) Recombination between the concatenated and native chromosomes is predicted to lead to the formation of deleterious and genetically unstable dicentric (or multicentric) chromosomes[22].

We have shown that *S. cerevisiae* can survive with two chromosomes and grow relatively well with four chromosomes, at least under laboratory conditions. Why does it have sixteen small chromosomes whereas *S. pombe* can get by with only three larger ones?

We consider three possibilities for phylogenetic distribution of n. First, the extant pattern of values of n in yeast could be a historical product of a WGD event(s) and/or spontaneous chromosome fusions and breaks[3]. *Saccharomyces* ($n=16$) and related genera arose as a WGD event relative to a large cluster of 'preduplication' yeasts[3–5] (mostly $n=8$). Second, a genome with 16 small chromosomes readily becomes aneuploid, allowing rapid and reversible adaptation to severe environmental changes[23]. However, aneuploidy can also result in deleterious imbalances[24,25]. With few large chromosomes, aneuploidy is more likely to be selected against, limiting potential for adaptation to environmental changes. Finally, sub-telomeres are enriched in 'contingency' genes that encode functions related to the cell wall and metabolism of different nutrients[26]. These genes are transcriptionally repressed under normal conditions but can be specifically expressed in appropriate environments or conditions[27]. A higher number of telomeres may allow more elaborate fine-tuning of these properties, improving the ability of a generalist species to adapt more rapidly to diverse environments and stresses.

We have efficiently generated a series of chromosome fusions, building a collection of strains with progressively fewer chromosomes, including $n=2$. Unexpectedly, yeast growth was robust, and mostly indistinguishable from normal karyotype strains, even with chromosomes up to four times the maximum size (excluding rDNA) of wild-type strains, and with greatly reduced numbers of telomeres and centromeres. The $n=2$ strain grew slightly more slowly than the wild type, perhaps owing to its large chromosomes or altered TPE. Meiosis was strongly affected in matings between strains with different numbers of chromosomes, suggesting that chromosome fusion and concomitant reduction in chromosome number suffices as a reproductive barrier. The strains described here may be used to probe aspects of meiotic recombination, replication origin timing, the role of 3D nuclear structure in transcriptional regulation in yeast, or recombination donor preference, to name a few. Reproductive isolation may also prove to be useful for future studies involving field release of engineered yeast or other cases in which genetic isolation is desirable.

The $n=2$ strain is reproductively isolated from the $n=16$ strain, but is it a new species? By classical biological species definitions (post-zygotic reproductive isolation)[28,29], it could qualify. However, by phylogenetic species definitions (a group of organisms sharing an ancestor), $n=2$ is not a new species because it has near-zero sequence divergence from $n=16$, and its morphology is similar[30]. Eight fusions suffice to cause practically complete reproductive isolation, which, with further neutral evolution, would allow such pairs of strains to accumulate diverse mutations over time, leading to speciation by any definition.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0374-x.

Received: 16 July 2017; Accepted: 25 June 2018;
Published online 1 August 2018.

1. Kandul, N. P., Lukhtanov, V. A. & Pierce, N. E. Karyotypic diversity and speciation in *Agrodiaetus* butterflies. *Evolution* **61**, 546–559 (2007).
2. Taylor, R. W. Ants with Attitude: Australian Jack-jumpers of the *Myrmecia pilosula* species complex, with descriptions of four new species (Hymenoptera: Formicidae: Myrmeciinae). *Zootaxa* **3911**, 493–520 (2015).
3. Gordon, J. L., Byrne, K. P. & Wolfe, K. H. Mechanisms of chromosome number evolution in yeast. *PLoS Genet.* **7**, e1002190 (2011).
4. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
5. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
6. IJdo, J. W., Baldini, A., Ward, D. C., Reeders, S. T. & Wells, R. A.; JW. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc. Natl Acad. Sci. USA* **88**, 9051–9055 (1991).
7. Ueda, Y. et al. Large-scale genome reorganization in *Saccharomyces cerevisiae* through combinatorial loss of mini-chromosomes. *J. Biosci. Bioeng.* **113**, 675–682 (2012).
8. Neurohr, G. et al. A midzone-based ruler adjusts chromosome compaction to anaphase spindle length. *Science* **332**, 465–468 (2011).
9. Titos, I., Ivanova, T. & Mendoza, M. Chromosome length and perinuclear attachment constrain resolution of DNA intertwines. *J. Cell Biol.* **206**, 719–733 (2014).
10. DiCarlo, J. E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR–Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
11. Schubert, I. & Oud, J. L. There is an upper limit of chromosome size for normal development of an organism. *Cell* **88**, 515–520 (1997).
12. Verdaasdonk, J. S. & Bloom, K. Centromeres: unique chromatin structures that drive chromosome segregation. *Nat. Rev. Mol. Cell Biol.* **12**, 320–332 (2011).
13. Winey, M. & Bloom, K. Mitotic spindle form and function. *Genetics* **190**, 1197–1224 (2012).
14. Shao, Y. et al. Creating a functional single chromosome yeast. *Nature* https://doi.org/10.1038/s41586-018-0382-x (2018).
15. Lynch, M. et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl Acad. Sci. USA* **105**, 9272–9277 (2008).
16. Wang, Y., Shirogane, T., Liu, D., Harper, J. W. & Elledge, S. J. Exit from exit: resetting the cell cycle through Amn1 inhibition of G protein signaling. *Cell* **112**, 697–709 (2003).
17. Baack, E., Melo, M. C., Rieseberg, L. H. & Ortiz-Barrientos, D. The origins of reproductive isolation in plants. *New Phytol.* **207**, 968–984 (2015).
18. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr. Biol.* **24**, 1153–1159 (2014).
19. Delneri, D. et al. Engineering evolution to study speciation in yeasts. *Nature* **422**, 68–72 (2003).
20. Liti, G., Barton, D. B. & Louis, E. J. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics* **174**, 839–850 (2006).
21. Leducq, J. B. et al. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nat. Microbiol.* **1**, 15003 (2016).
22. Haber, J. E., Thorburn, P. C. & Rogers, D. Meiotic and mitotic behavior of dicentric chromosomes in *Saccharomyces cerevisiae*. *Genetics* **106**, 185–205 (1984).
23. Yona, A. H. et al. Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl Acad. Sci. USA* **109**, 21010–21015 (2012).
24. Sheltzer, J. M. et al. Aneuploidy drives genomic instability in yeast. *Science* **333**, 1026–1030 (2011).
25. Bonney, M. E., Moriya, H. & Amon, A. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. *Genes Dev.* **29**, 898–903 (2015).
26. Brown, C. A., Murray, A. W. & Verstrepen, K. J. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* **20**, 895–903 (2010).
27. Ai, W., Bertram, P. G., Tsang, C. K., Chan, T. F. & Zheng, X. F. Regulation of subtelomeric silencing during stress response. *Mol. Cell* **10**, 1295–1305 (2002).
28. Mayr, E. *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942).
29. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
30. De Queiroz, K. Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886 (2007).

Acknowledgements We thank L. A. Mitchell, N. Agmon, and D. M. Truong for discussion and comments on this manuscript; L. A. Vale Silva and A. Hochwagen for sharing strains, advice and reagents; The NYU Langone Health Genome Technology Center, especially A. Heguy and P. Zappile, for deep sequencing libraries; M. S. Hogan and M. T. Maurano for help with the NextSeq 500 sequencer; Z. Kuang, X. Wang and Z. Tang for advice on bioinformatics analysis; and M. Delarue and L. Holt for help and access to spinning disk confocal microscopy. This work was supported by NSF grant MCB-1616111 to J.D.B.

Reviewer information *Nature* thanks G. Liti, K. Wolfe and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions J.L., B.P.C. and J.D.B conceived the project; J.L. and J.D.B. designed experiments, analysed results and wrote the manuscript; J.L. performed experiments; X.S. performed bioinformatics analysis; all authors read and commented on the manuscript.

Competing interests J.D.B is a founder and Director of Neochromosome, Inc., the Center of Excellence for Engineering Biology, and CDI Labs, Inc. and serves on the Scientific Advisory Board of Modern Meadow, Inc., Recombinetics, Inc., and Sample6, Inc. All other authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0374-x.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0374-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.D.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Strains. All strains that we constructed in this study are listed in Extended Data Table 2. They are all derived from BY4741 (*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0*) by transformation events. For sporulation efficiency measurement experiments, the SK1 strain (*MATα ho::LYS2, lys2, ura3, leu2::hisG, his3::hisG, trp1::hisG*) was used to mate with strains of different *n*. For the tetrad dissection experiments, BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was used to mate to strains with different values of *n*. 17 loxPsym sites were added at the sites of centromere deletion and telomere fusion anticipating future genome rearrangement experiments.

Calculation for potential paths to fuse 16 chromosomes into one chromosome. The equation for this calculation is as follows: $(32 \times 30 \times 28 \times \ldots \times 2 \times 16)/2 = 1.097 \times 10^{19}$, if we take into consideration final centromere choice (16 conditions). In addition, we count the same configuration twice by arbitrarily defining left arm and right arm. To give a simple example to explain this point, suppose there are only two chromosomes: chr I and chr II, the final fusion chromosome chr IL-chr IR-chr IIL-chr IIR is the same as chr IIR-chr IIL-chr IR-chr IL, thus we divide by 2 to account for this.

CRISPR–Cas9 method. The CRISPR–Cas9 method for genome editing in *S. cerevisiae* dramatically stimulates homologous recombination (HR) by co-expressing Cas9 and site-specific guide RNAs, together with transient linear PCR products as HR donors[10]. Here, we optimized this method to produce chromosome fusions. To fuse chrs I and III, for example, we targeted breaks adjacent to the chr III right telomere and chr I left telomere, and near CEN1 by Cas9 co-expression with three specific gRNAs (Fig. 2a). We synthesized two donor molecules, the first encoding homology to bridge the termini of these chromosomes, and the second a fragment of DNA spanning CEN1 to delete it. Through successful HR at both sites, we generated a compound chromosome III-I, with deletion of two telomeres, chr IIIR_tel and chr IL_tel, and one centromere, CEN1 in a single shot. The choice of gRNA telomere target site was chosen to optimize targeting specificity while minimizing deletion of sub-telomeric sequences to preserve genome content. At each stage, we picked plasmid-transformed colonies and then screened by PCR for a) presence of appropriate telomere-telomere junctions and b) deletion of appropriate centromeric DNA (Extended Data Fig. 1a).

The CRISPR–Cas9 method was performed as described[10,31]. In addition, another gRNA acceptor vector was constructed from gRNA-ura-HYB (Addgene plasmid #64330) by substituting the *URA3* gRNA with a NotI restriction site. pRS426 gRNA acceptor vectors have two gRNA insertion sites, one defined by a NotI restriction site and another one by a HindIII restriction site. By using these two gRNAs acceptor vectors, up to 3 gRNAs could be co-expressed in yeast cells when co-transformed. To insert 20 nt gRNAs, Gibson assembly[32] was used to assemble a restriction enzyme digested (for example, NotI or HindIII) and gel purified vector together with a 60 bp double-stranded oligonucleotides, which consists of a 20nt gRNA, and homologous sequences to the left and right of vectors (for example, for the NotI vector, 5′-GCAGTGAAAGATAAATGATC-20nt gRNA -GTTTTAGAGCTAGAAATAGC-3′; for the HindIII vector, 5′- CTGGGA GCTGCGATTGGCAG-20nt gRNA-GTTTTAGAGCTAGAAATAGC-3′). Single-stranded oligonucleotides were ordered from Integrated DNA Technologies and annealed[10] by first denaturing the mixture at 95 °C for 5 min and then cooling to 10 °C with a ramp of 0.17° s^{-1}. The donor DNAs linking two different chromosome arms together are made by two-step fusion PCR[33] amplified from the wild type yeast genome or they were synthesized. These 'linker donors' usually have ~ 400bp sequences homologous to each chromosome arm (sometimes shorter sequences were used to avoid repetitive terminal sequences that would not target uniquely). The same rule applies for the centromere deletion donors, ~ 400bp flanking centromere regions were designed on each side. The centromere deletion donors were made by Polymerase Chain Assembly[34,35]. Donor DNA amplicons were about 800 bp long. The CRISPR–Cas9 method was performed stepwise: 1. We transformed a Cas9 expressing plasmid. 2. We co-transformed 50 ng gRNA expression plasmids and ~400 ng PCR amplicon donor DNAs into the strain already expressing Cas9.

Both pRS426 and pRS42H carrying 3 gRNAs in total were used, because we found that using 3 gRNAs gave a higher yield of correct colonies than only using 2 gRNAs. Then after co-transformation, yeast cells were plated on SC–Ura–Leu (Synthetic Complete medium lacking leucine and uracil) with 300 μg/ml hygromycin B plates for more than 2 days. Yeast cells were usually grown for 4 h in YPD after transformation to allow expression of hygromycin resistance, before plating on selective plates with 300 μg/ml hygromycin B. The yeast clones with correct fusion chromosome configurations were verified by PCR and pulse field electrophoresis. Sometimes the SC–Ura–Leu + Hyg plate selection is too stringent in that very few transformant colonies were obtained. In that case, SC–Ura–Leu plates were used.

PCR verification of fusion chromosomes. To verify the fusion of two chromosomes, primers were designed to bind outside of homology sequences to the linker donors. Only clones bearing successfully reorganized fusion karyotypes generated appropriately sized PCR amplicons, which were absent from failed clones or wild type control colonies. To verify the deletion of a centromere, primers were designed to bind sequences flanking centromere regions. Typically, the successful deletion of centromeres produced amplicons ~120 bp shorter than wild type control samples. All verification primer information is included in Extended Data Table 5.

Pulsed-field gel electrophoresis. For $n = 16$ to $n = 9$ strains (Fig. 2b), yeast chromosomes plugs were prepared and separated by clamped homogeneous electric field (CHEF) gel electrophoresis using the CHEF-DR III Pulsed-Field Electrophoresis System (Bio-Rad), as previously described[36]. The CHEF gel running condition was 6 V/cm, switch time: 60 s to 120 s, run time: 24 h, 14 °C, with a 0.5 × Tris-Borate-EDTA buffer and a 1% gel with low melting point agarose. For $n = 9$ to $n = 2$ strains (Figs. 2c and 2d), a 7 ml culture of yeast was grown over 48 h or until it reached stationary phase. Then yeast cell pellets were collected by centrifugation and zymolyase and agarose were added proportion to cell pellet weight. For 60 mg of cell pellet, 24 μl 25 mg/ml zymolyase 20T in 10 mM KPO4 (pH 7.5) and 540 μl 0.5% low melting point agarose in 100 mM EDTA (pH 7.5) were added. Low melting point agarose was fully dissolved in the EDTA solution by microwave heating and then kept at 42 °C. Cell pellets were resuspended with zymolyase and agarose solutions by pipetting with a wide bore pipette and transferred to BioRad molds (BioRad No.1703713). The molds were cooled in a 4 °C cold room for 30 m. The plugs were released from the molds, and added to 1 ml 500 mM EDTA, 10 mM Tris, pH 7.5 and incubated at 37 °C overnight. 5% sarcosyl and 5 mg/ml proteinase K in 500 mM EDTA pH 7.5 was added to the plug the next day and incubated at 50 °C overnight. The plugs were washed with 1 ml 2mM Tris-1mM EDTA buffer four times, for 1 h each. Such plugs could be used immediately or stored at 4 °C for at least a year. To separate 1 Mb to 3 Mb chromosomes, we used *H. wingei* chromosomal DNA as a marker. *H. wingei* has 7 chromosomes, varying from 1.05 Mb to 3.13 Mb. For $n = 9$ to $n = 4$ strain (Fig. 2c), the CHEF gel electrophoresis condition was that recommended for *H. wingei* chromosomes (BioRad No.170-3667). To separate chromosomes longer than 3 Mb, we adopted a different electrophoretic protocol, using *S. pombe* chromosomal DNA as a marker. For $n = 4$ to $n = 2$ strain (Fig. 2d), the CHEF gel electrophoresis condition was that recommended for *S. pombe* chromosomes (BioRad No.170-3633).

Calculation of maximum arm length. The condensation of a fusion chromosome, chr IV-XII, in budding yeast has been studied by labelling *TRP1* with red fluorescence and *LYS4* with green fluorescence[8]. In mother cells, the distance between *TRP1* and *LYS4* was ~0.4 μM during anaphase. With ~470 kb between these two loci, extrapolating from this distance predicts that each 1 Mb of condensed chromosome during anaphase will extend ~0.85 μm (0.4 μm/0.4686 Mb). Anaphase spindle axis length can extend up to 10 μm[13], suggesting the maximum distance a chromosome arm could extend is 5 μm. The maximum arm length limit, therefore, should be about 5.9 Mb (5 μm/0.85 μm per 1Mb).

Serial dilution assays. Yeast strains were grown from single colonies in liquid YPD culture until they reached the stationary phase at 30 °C with rotation. Then culture was diluted to $A_{600} = 0.01$, and serially diluted (1:10) in water and plated on different media. YPG plates were prepared by adding 3% glycerol as carbon source to yeast extract peptone. All other compounds (HU, MMS, benomyl) were added to YPD medium. Plates were incubated at 30 °C for 2 days except for YPG plates, HU plates and MMS plates, which were incubated for 3 days.

Growth curve measurement. Yeast strains were grown from single colonies in liquid YPD culture until they reached stationary phase at 30 °C with rotation. Then culture was diluted to $A_{600} = 0.05$ and grown in 96 well plates with 100 μl YPD or YPD + 0.2 M hydroxyurea for 48 h. Every 10 m, the plate would be shaken and a measurement of OD600 was taken by a BioTek Eon microplate spectrophotometer. Doubling time was measured by calculating the slope of the growth curve at exponential stage.

Competitive growth assays. Competitive growth assays were carried out as described[37]. The GFP or dTomato cassette was integrated into the *HO* locus through selection of nourseothricin-resistant cells. The GFP-labelled BY4741 was co-cultured with dTomato-labelled BY4741 or $n = 4$ in 1:1 ratio, while the GFP-labelled BY4741 was mixed with dTomato-labelled $n = 2$ in 1:2.5 ratio in YPD. The next day, 30,000 total cells were sorted on a Sony SH800S Cell Sorter as T0. Cells were diluted 200-fold in fresh YPD medium and sorted every 24 h three times. Flowjo v.10 was used to analyse the data.

Genome sequencing. Whole-genome DNA samples for sequencing were prepared using a Norgen Biotek fungi/yeast genomic DNA isolation kit (Cat No. 27300). Whole-genome shotgun libraries were made on a Beckman FXP automation workstation and prepared as follows: 500 ng genomic DNA, as input, was amplified by 2 cycles of PCR and sheared to 500 bp on a Covaris LE220. A library was prepared using KAPA High Throughput Library Preparation Kit (KK8234), and sequenced as 75 bp paired-end reads on an Illumina NextSeq 500. All raw reads were trimmed to remove adaptor sequence using Trimmomatic, and subsequently mapped to UCSC sacCer2 reference genome (S288C_reference_genome_R61-1-1_20080605) from Saccharomyces Genome Database using BWA-MEM standard options. BWA, Picard, GATK and SAMtools software were applied to align reads to transcriptome reference and call SNPs/Indels. The filtering criteria for SNPs using GATK were as

follows: QD <2.0; FS >60.0; MQ <40.0; MQRankSum <-12.5; ReadPosRankSum <-8.0; SOR >4.0. The filtering criteria for Indels using GATK are as follows: QD <2.0; FS >200.0; ReadPosRankSum <-20.0; SOR >10.0. The variants common to the fused chromosome and the laboratory stock wild-type strain BY4741 were removed and considered as starting spontaneous mutations. The remaining variants were manually curated, by browsing through the bam files in IGV. Native 2-μm plasmids were absent from those sequenced fusion chromosome strains $n = 12$ to $n = 2$, according to the WGS results.

RNA sequencing. For each strain, three independent colonies were grown in YPD liquid medium at 30 °C with rotation to saturation. The next day, an overnight culture was diluted to $A_{600} = 0.1$ and regrown. Cells were harvested at $A_{600} = 0.4$–0.6. Total RNA was extracted from three independent log-phase cultures using a QIAGEN RNeasy mini kit (Cat No.74106). The library was prepared with 500 ng total RNA as input, using a TruSeq RNA sample Preparation v2 kit (set A RS-122-2001 and set B RS-122-2002) with 13 cycles of amplification. The library was sequenced as 150 bp single end reads on an Illumina NextSeq 500. Reads were mapped to S288C reference genome (sacCer2) and differential gene expression analysis was performed with TopHat and Cuffdiff according to a standard pipeline[38]. Briefly, trimmed reads were mapped to the reference genome using TopHat, and aligned reads with more than 2 mismatches were excluded from the downstream analysis. RPKM were calculated using cufflinks and differentially expressed genes were analysed using cuffdiff. Genes that have been deleted in the fusion chromosomes were excluded in the analysis. The thresholds we used for differentially expressed genes are $P < 10^{-5}$ and |fold change|>2.

Definition of Isogenic strain. Isogenic, as used here, refers to genomic regions not deleted together with centromeres and telomeres. The deleted regions contain a maximum of 21 verified genes ($n = 2$ strain) as defined by SGD, and a maximum of 1.6% of the genome in the most extreme case, that is, comparing $n = 2$ and $n = 16$ strains, and consist primarily of the centromeres and telomeres themselves, and some adjacent repetitive subtelomeric elements, including X elements, Y' elements, and highly repeated *PAU* and *COS* genes.

Mating type switching. Mating type switch of fusion chromosome strains were carried out by transforming a CEN plasmid, which expresses endonuclease HO with its endogenous promoter, as described[39].

DAPI staining of fixed yeast cells. 10^7 cells were spun down from an overnight culture. Pelleted cells were fixed by re-suspending in 70% EtOH and incubated for 2 h in room temperature. Cells were spun down, washed with water, re-suspended in 500 μl RNase A solution (2 mg/ml RNase A (Qiagen, 19101), 50 mM Tris pH 7.5 and 15 mM NaCl) and incubated for 2 h at 37 °C. Proteinase K (Invitrogen, 25530049) was added to a final concentration 1 mg/ml and incubated for 45 min at 37 °C. Cells were collected by centrifugation, washed with 1 ml of 50 mM Tris, and stored in 1ml of 50 mM Tris at 4 °C.

100–200 μl of fixed cells were pelleted by centrifugation, stained by re-suspending in 100 μl of DAPI solution (300 nM DAPI in 1× PBS) and incubated for 15 min. Cells were spun down and pellet was washed with 1× PBS, spun down again and re-suspended in 1× PBS for imaging.

Images were collected with NIS-Elements acquisition software, with 100× oil objective lens using Nikon Eclipse Ti microscope in both DAPI channel and bright field channel. Images were cropped in NIH ImageJ software with Bio-Formats plugin.

Sporulation and tetrad dissection. For BY4743 diploid background, a diploid clone was inoculated in YPD at 30 °C with rotation overnight. The next day culture was diluted in YPD medium to grow at 30 °C for at least two generations to $A_{600} = 4$–8. The cells were pelleted by centrifugation (2,000*g* for 2 m) and washed three times with reagent-grade water. Cell pellets were re-suspended in sporulation medium (1% potassium acetate and 0.005% zinc acetate) with 0.1% (w/v) yeast extract and amino acid supplements (0.3 mM histidine, 2 mM leucine and 0.2 mM uracil), at a final $A_{600} = 1.0$, and incubated at 25 °C with rotation for 3–15 days before tetrad dissection. To digest asci, 50 μl cell cultures were pelleted by centrifugation (3,600 rpm × 3 m), re-suspended in 25 μl 0.5 mg/ml 20T zymolyase in 1 M sorbitol at 37 °C for 6–7 m, and diluted with 300 μl 1M sorbitol. Tetrads were dissected on a SINGER Instruments dissection microscope and germinated on YPD plates at 30 °C for 2 days. yJL346 was used in this experiment as the $n = 8$ strain rather than yJL342.

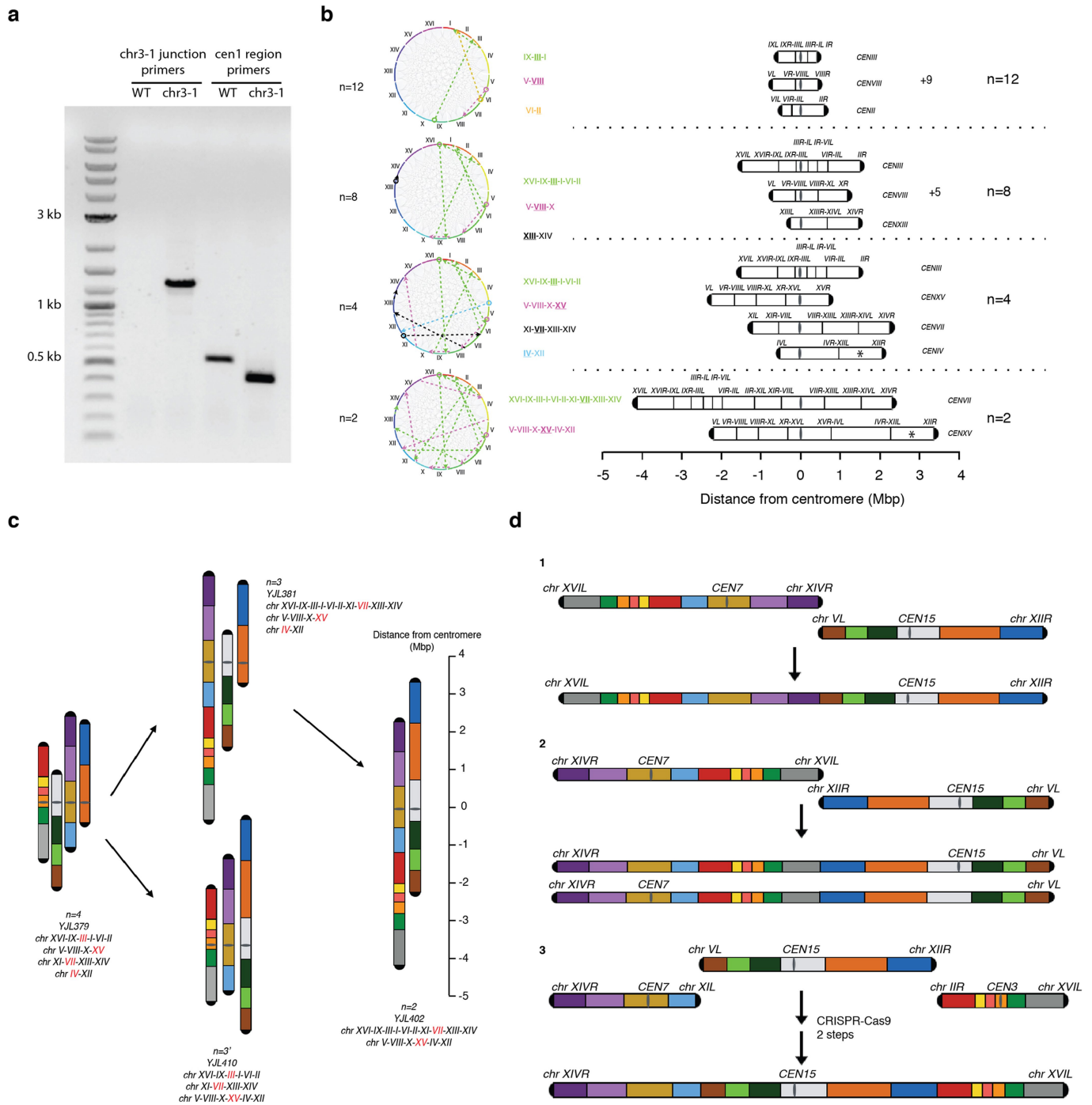
For crosses of the hybrid SK1 × BY4741 background, diploid strains were patched onto a YPG plate from −80 °C glycerol stocks and grown at 30 °C overnight (O/N), then patched onto a YPD plate containing 4% glucose and grown at 30 °C O/N, inoculated into YPD liquid culture and grown at 25 °C O/N, diluted to $A_{600} = 0.8$ and then grown in BYTA pre-sporulation medium (buffered yeast extract tryptone acetate, 1% yeast extract, 2% bacto tryptone, 1% potassium acetate and 50mM potassium phtalate) at 30 °C with rotation for ~16.5 h. Cell pellets were collected by centrifugation (3000 rpm × 5 min), washed twice with 25 ml water, re-suspended to $A_{600} = 2.0$ in sporulation medium (0.3% potassium acetate, 5% acetic acid, pH ~ 6.5, with 0.3mM histidine, 2 mM leucine and 0.2 mM uracil), incubated at 30 °C with shaking. The sporulation efficiency of each strain was measured at 48 h on a ZEISS phase contrast microscope under 60× magnification. For each strain, > 100 cells were counted to obtain the sporulation efficiency. We scored asci with 0 spore and cells that did not go through meiosis in the same category (asci with 0 spore). We calculated sporulation efficiency by dividing the number of asci with at least 1 spore to total cells. yJL346 was used in this experiment as $n = 8$ strain rather than yJL342.

Statistics and reproducibility. Figures 2b–d, 3a and 4c, d and Extended Data Figs. 1a, 2a and 5a were repeated twice with similar results. Extended Data Fig. 2b was done with biological quadruplicates.

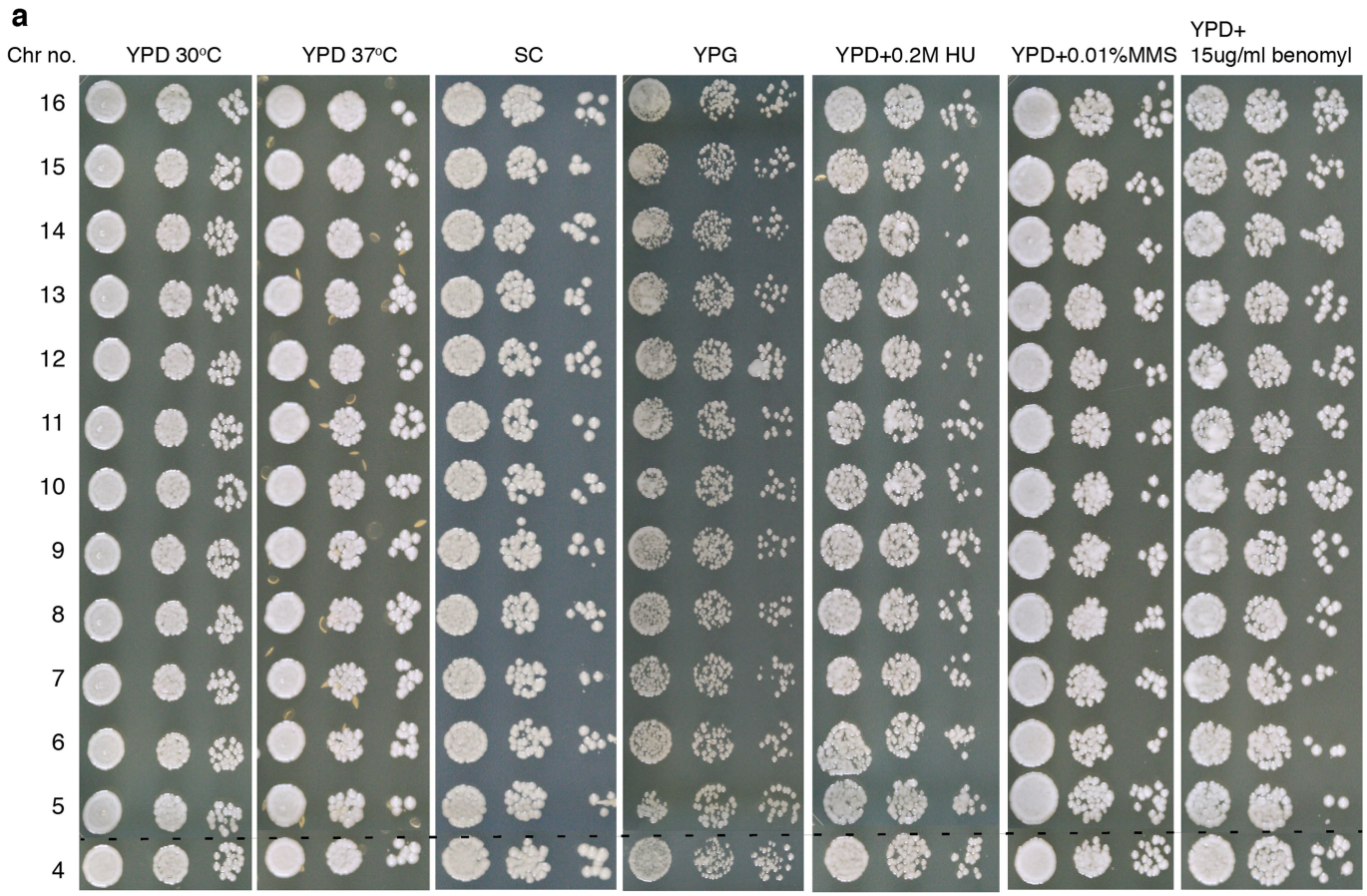
Code availability. All codes used in this study are available at https://github.com/sunnysun515/fusion_yeast_chromosomes.

Data availability. We submitted DNA sequences and RNA sequencing data to NCBI BioProject PRJNA471518. All other data are available from the corresponding author upon reasonable request.

31. Neta Agmon, J. T. et al. Human to yeast pathway transplantation: cross-species dissection of the adenine de novo pathway regulatory node. Preprint at https://www.biorxiv.org/content/early/2017/06/14/147579 (2017).
32. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
33. Yon, J. & Fried, M. Precise gene fusion by PCR. *Nucleic Acids Res.* **17**, 4895 (1989).
34. Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
35. Richardson, S. M., Wheelan, S. J., Yarrington, R. M. & Boeke, J. D. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**, 550–556 (2006).
36. Mitchell, L. A. et al. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355**, eaaf4831 (2017).
37. Breslow, D. K. et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5**, 711–718 (2008).
38. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* **7**, 562–578 (2012).
39. Wu, Y. et al. Bug mapping and fitness testing of chemically synthesized chromosome X. *Science* **355**, eaaf4706 (2017).
40. Ellahi, A., Thurtle, D. M. & Rine, J. The chromatin and transcriptional landscape of native *Saccharomyces cerevisiae* telomeres and subtelomeric domains. *Genetics* **200**, 505–521 (2015).



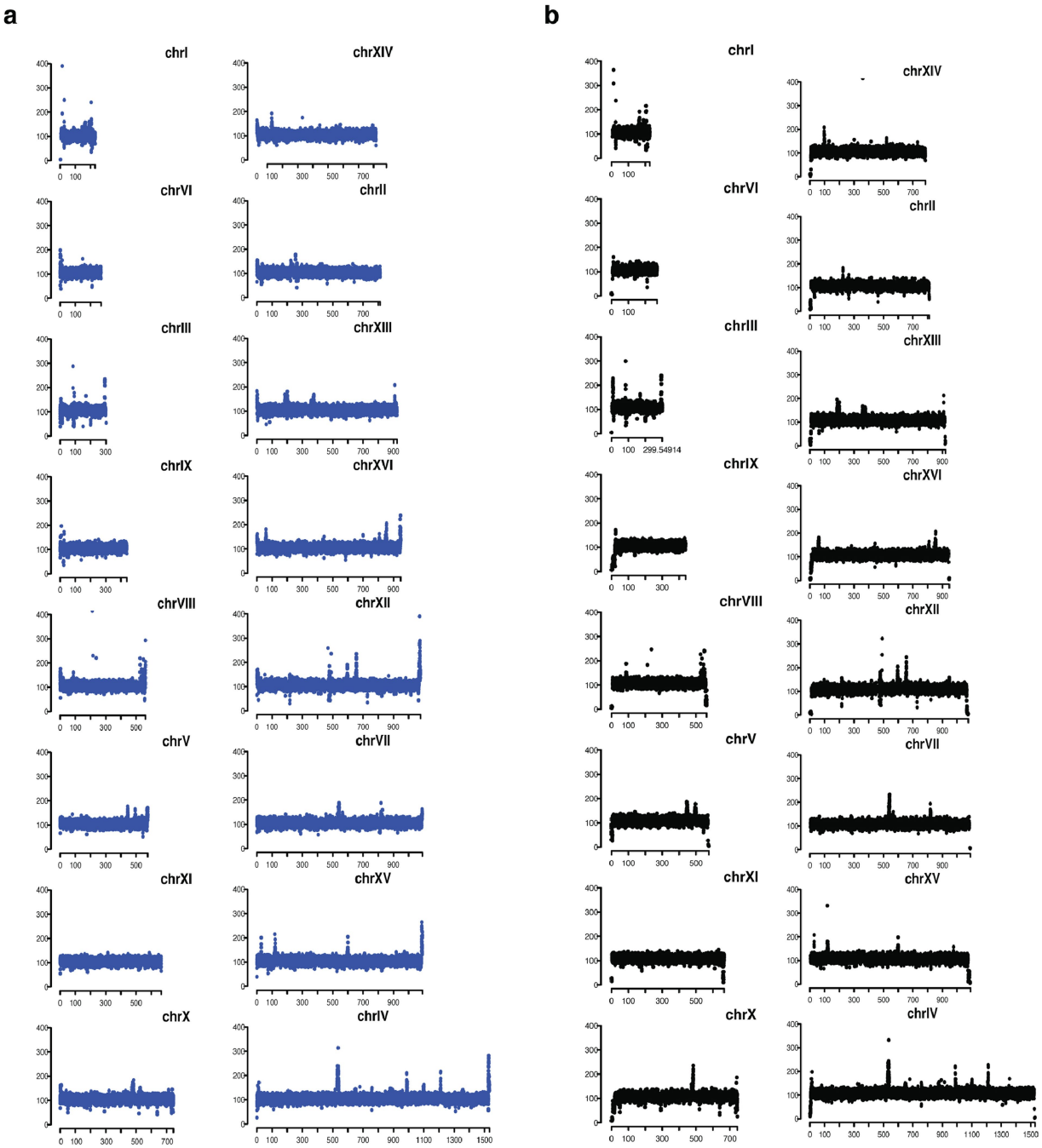
Extended Data Fig. 1 | Fusion chromosome paths and characterization. **a**, PCR verification of fusion chromosomes. Two pairs of primers are used to verify the fusion of two chromosomes, here presenting the fusion of chromosomes I and III. Only with successful fusion would an amplicon show in fusion chromosome junction PCR. In addition, the centromere region PCR amplicon is shorter, owing to the deletion of a centromere. Marker: 2-log DNA marker. **b**, Circos diagrams for different values of *n* are shown in the upper panel, with 16 chromosomes laid out in circles in a clockwise manner. Each grey line connecting each pair of chromosomes represents a possible intertelomeric link. The dashed coloured lines represent the path we chose. (Open circle indicates the starting chromosome, arrows show direction of fusion chromosomes.) Each fusion chromosome is labelled in the same colour, providing detailed information on fusion chromosome paths. The underlined chromosome has the active centromere for the fusion chromosome. Unchanged chromosomes are not shown, but the number of unchanged chromosome is indicated after +, for example, +9 means nine unchanged chromosomes. Fusion chromosome lengths are shown below. Please note that the length of the rDNA array (normally 1–2 Mb)is omitted here. *Chromosome containing rDNA array. In addition, telomeric ends are clearly labelled with the original chromosomes; L/R stands for left/right telomere. Active centromeres are written on the right of each fusion chromosome diagram. When we deleted CEN15 and made 4 bp of mutations in gRNA sites in BY4741 to fuse chromosomes X and XV together, the fusion strain grew more slowly than the wild type. The growth defect could be due to altered expression of the gene neighbouring CEN15 (YOR001W; *RRP6*). For this reason, CEN15 was maintained in the remaining centromere. **c**, Fusion chromosome paths from $n=4$ to $n=2$. Red lettering indicates the chromosome that has an active centromere in the compound chromosome. **d**, Multiple strategies that we attempted to construct an $n=1$ strain. We tried changing chromosome arm length (strategy 3), centromere position (strategy 2), and which telomeres remained (strategy 1 and 2). In strategy 2, we attempted both versions of $n=1$ strains, keeping either CEN7 or CEN15 as the active centromere. In addition, we also attempted the strategy 1 in both *SIR2*$^{+}$ and *sir2*Δ backgrounds.



c

Average doubling time (min)	BY4741	n=4	n=2
YPD	135	141	182
YPD+0.2M HU	222	208	278

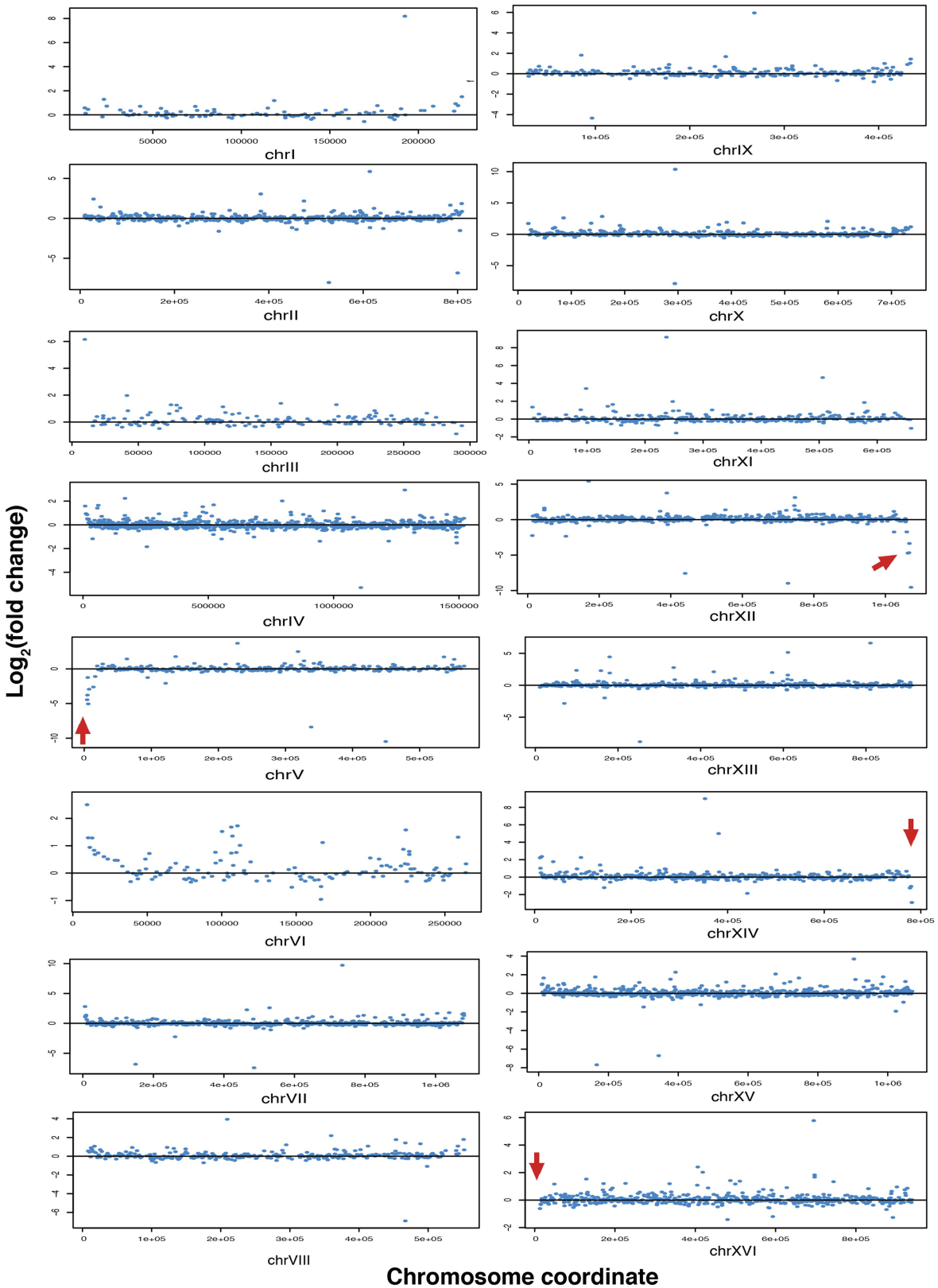
Extended Data Fig. 2 | Growth fitness assays. a, Serial dilution assays in seven different conditions with $n = 16$ through $n = 4$ strains. **b**, Growth curve for BY4741, the $n = 4$ strain and the $n = 2$ strain in YPD medium at 30 °C. **c**, Doubling time calculation for these three strains. Each experimental group was tested in biological quadruplicate.



c

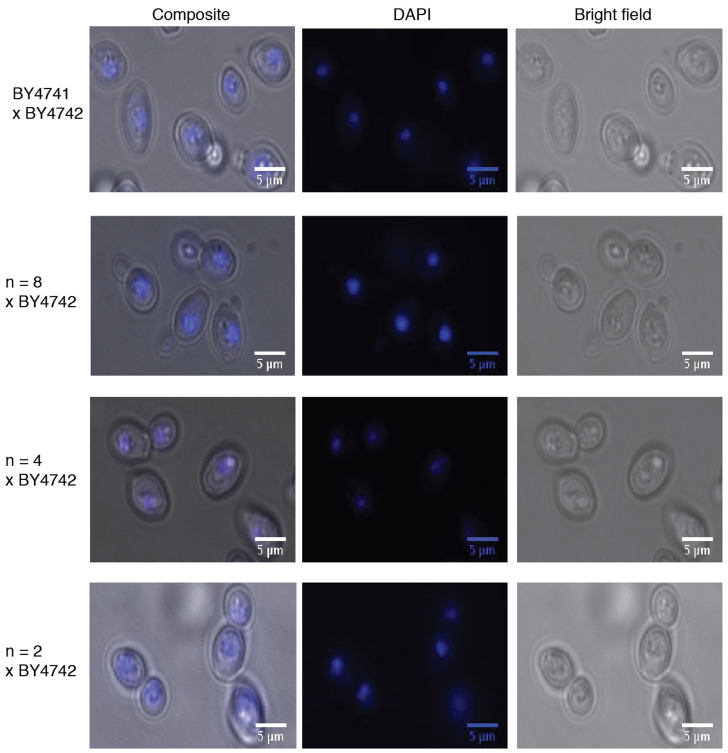
	Detected number of SNPs	Expected number of SNPs
n=12	3	1.1
n=8	4	2.2
n=4	9	3.4
n=2	11	3.9

Extended Data Fig. 3 | Whole-genome coverage maps and SNPs comparison. a, Whole-genome coverage map of BY4741. Chromosomes are ranged according to their size. *x*-axis shows chromosome coordinate, and *y*-axis shows the coverage of reads that map to the reference genome of S288C. **b**, Whole-genome coverage map of $n = 2$ strain. For comparison, we mapped the reads back to the S288C genome. The main difference is telomere reads. Owing to deletion of telomeres in $n = 2$, either no reads or only few reads align back to fused telomere ends. **c**, Each round of CRISPR–Cas9 experiments takes at least 70 generations. The nucleotide mutation rate is 0.33×10^{-9} per base per generation[15]. If we multiply the number of rounds of CRISPR–Cas9 $(16 - n) \times 70$ generations × yeast genome size (bp) $12 \times 10^6 \times 0.33 \times 10^{-9}$, it gives the expected number of SNPs, as indicated in the table.

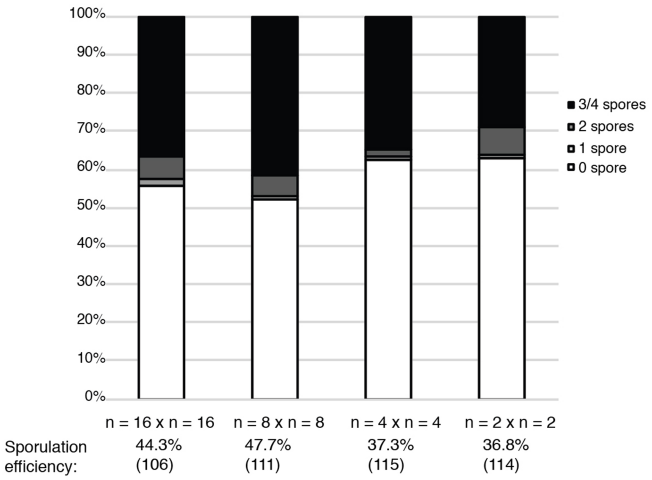


Extended Data Fig. 4 | Chromosome plots of gene expression change in $n = 2$ vs $n = 16$ strains. y-axis, $\log_2$ (fold change of gene expression ($n = 2/n = 16$)); x-axis, chromosome coordinates. Red arrows point to the four remaining telomeres in $n = 2$ strains.

a



b



c

crosses	fractions of 4-viable %	fractions of 3-viable%	fractions of 2 viable%	fractions of 1 viable%	fractions of 0 viable%	total dissected
n = 16 x BY4742	82.14	17.86	0	0	0	28
n = 14 x BY4742	7.14	10.71	21.43	32.14	28.57	28
n = 12 x BY4742	0	0	0	28.57	71.43	28
n = 8 x BY4742	0	0	0	0	100	16
n = 16 x n = 16	90.63	6.25	3.13	0	0	32
n = 8 x n = 8	93.75	6.25	0	0	0	32
n = 4 x n = 4	93.75	3.13	3.13	0	0	32
n = 2 x n = 2	80.65	19.35	0	0	0	31

Extended Data Fig. 5 | DAPI staining of nucleus and sporulation efficiency for diploids. a, DAPI staining of nuclei in diploids. **b**, Sporulation efficiencies for $2n = 32$, $2n = 16$, $2n = 8$ and $2n = 4$ homotypic diploids. *y*-axis: percentage of asci that have 0–4 spores. *x*-axis: different diploid strains. Sporulation efficiency was measured after shifting diploid strains in the sporulation medium for 10 days in room temperature. **c**, Fractions of asci with 0–4 viable spores in heterotypic and homotypic crosses.

Extended Data Table 1 | gRNA sequences and fusion efficiency of each fusion step

Fusion step	Deletion	20 nt gRNA (5'->3')	Deletion coordinate	Number of colonies	PCR verification
1	*chr III_right_ telomere* *chr I_left_ telomere* CEN1	GAGCTACTATCTTTGTCGGG CTCAATGTACGCGCCAGGCA AAGAAAGTTATATGTGTGAC	299737 - 316617 1 - 6770 151467 - 151578	10*	4
2	*chr IX_right_ telomere* *chr III_left_ telomere* CEN9	GTTGAGAGACAGGATGGTTA CCATTTGGAGTCTGCTCGGC TGAATAAGTTGAAGGACAAC	438945 - 439885 1 - 3582 355626 - 355853	NA	NA
3	*chr V_right_ telomere* *chr VIII_left_ telomere* CEN5	TGCTAATCGTCTCTGCGGTT CCATTTTTTATCCTTATGCT TTTAGTTGAAACGCCAACAG	569244 - 576869 1 - 6182 151813 - 152187	1	1
4	*chr VI_right_ telomere* *chr II_left_ telomere* CEN6	GAACTGTGCATCCACTCGTT GGGTTAACTACGCTATAGAC CTAAAACTGTCTTTTCGTGT	269599-270148 1-8703 148507-148624	3*	2
5	*chr XIII_right_ telomere* *chr XIV_left_ telomere* CEN14	CTGATAAAAGCAGGTGCCCT TCTGATCGGTCATACGTACA CTTCATCAAGATCGGGGAGC	916969-924429 1-7493 628758-628876	1*	1
6	*chr XVI_right_ telomere* *chr IX_left_ telomere* CEN16	TGGTGTTATATAGTGGCACC TTTGCCACCACCTGGGCGGG TTAGAATTACGACAACATAA	941911-948062 1-20857 555959-556070	1	1
7	*chr VIII_right_ telomere* *chr X_left_ telomere* CEN10	CATCCGTGTGCGTATGCCAT TAGTGGCTCGCACTCATGCG TTGTTATACACAACGCGTCT	556465-562643 1-7960 436301-436421	10	2
8	*chr I_right_ telomere* *chr VI_left_ telomere* CEN2	ACCTGCGCGGCGCGGCGGTT CTTTGTATGAGGGTACATCA CTTTGCGTGGTCTAGTGCAT	227349-230208 1-6185 238211-238325	48*	2
9	*chr VII*_right_ telomere *chr XIII*_left_ telomere CEN13	CTTTTACAACAACCGCCATA ATAGCGGCGGTACGTTTGTT CACTATTTATTTTACTATGG	1083059-1090947 1-8585 268036-268410	20*	17
10	*chr IV*_right_ telomere *chr XII*_left_ telomere CEN12	TTGTGGTAGCAACACTATCA GTTGTGTTGAGGTACTGTGT AGGAGAAAACTTGTAGTACG	1524527-1531919 1-11385 150818-150946	8	2
11	*chr XI*_right_ telomere *chr VII*_left_ telomere CEN11	CTGTTTAGACACTTGCGTCA AAAAGGATCTATCTCCCGCT TATTTAGTATTGGACCATTG	664993-666454 1-4700 439774-439891	16	9
12	*chr X*_right_ telomere *chr XV*_left_ telomere CEN8	CATCCAGATCGGAAACGCTA AGGTAAGGATACGGGGATAT TATTATACTAAATCGTTTTG	743832-745742 1-5056 105580-105700	2	1
13	*chr II*_right_ telomere *chr XI*_left_ telomere CEN3	CGTAACGTTGGTTCACATTT AAATGAAGAAGTGCCATGGG TAAGCGGAAGGGGAAGGGTT	811675-813178 1-2951 114487-114489	5*	1
14	*chr XV*_right_ telomere *chr IV*_left_ telomere CEN4	CTGCATCGTCTCGCTTTGCC CTTGCAACATCGCGACAGGT TCGGCATTTTGGCCGCTCCT	1073916-1091289 1-5566 449709-449819	135*	2 out of 95

*Selection plates are SC–Ura–Leu, otherwise are SC–Ura–Leu+Hyg. The same CRISPR–Cas9 experiment as fusion step 14 was also carried out in an $n = 15$ strain background. As a result, 71 colonies grew and 1 out of 71 passed the PCR verification, suggesting that the efficiency does not change much as n drops.

Extended Data Table 2 | Summary table of fusion chromosome strains

1	yJL373	Chr03R was fused to chr01L, with cen1 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 III:I (cen3)*	15
2	yJL374	Chr09R was fused to chr03L, with cen9 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 IX:III:I (cen3)*	14
3	yJL282	Chr05R was fused to chr08L, with cen5 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 IX:III:I (cen3) V:VIII(cen8)*	13
4	yJL303	Chr06R was fused to chr02L, with cen6 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 IX:III:I (cen3) V:VIII(cen8) VI:II(cen2)*	12
5	yJL310	Chr13R was fused to chr14L, with cen14 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 IX:III:I (cen3) V:VIII(cen8) VI:II(cen2) XIII:XIV(cen13)*	11
6	yJL320	Chr16R was fused to chr09L, with cen16 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I (cen3) V:VIII(cen8) VI:II(cen2) XIII:XIV(cen13)*	10
7	yJL336	Chr08R was fused to chr10L, with cen10 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I (cen3) V:VIII:X(cen8) VI:II(cen2) XIII:XIV(cen13)*	9
8	yJL342	Chr01R was fused to chr06L, with cen2 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II (cen3) V:VIII:X(cen8) XIII:XIV(cen13)*	8
8	yJL346	Chr07R was fused to chr13L, with cen13 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I (cen3) VI:II(cen2) V:VIII:X(cen8) VII:XIII:XIV(cen7)*	8
9	yJL358	Chr07R was fused to chr13L, with cen13 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II (cen3) V:VIII:X(cen8) VII:XIII:XIV(cen7)*	7
10	yJL369	Chr04R was fused to chr12L, with cen12 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II (cen3) IV:XII(cen4) V:VIII:X(cen8) VII:XIII:XIV(cen7)*	6
11	yJL375	Chr11R was fused to chr07L, with cen11 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II (cen3) IV:XII(cen4) V:VIII:X(cen8) XI:VII:XIII:XIV(cen7)*	5
12	yJL379	Chr10R was fused to chr15L, with cen8 deleted	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II (cen3) IV:XII(cen4) V:VIII:X:XV(cen15) XI:VII:XIII:XIV(cen7)*	4
13	yJL381	Chr02R was fused to chr11L, with cen3 inactived	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II:XI:VII:XIII:XIV (cen7) IV:XII(cen4) V:VIII:X:XV(cen15)*	3
13	yJL410	Chr15R was fused to chr04L, with cen4 deleted with parental strain: yJL379	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 V:VIII:X:XV:IV:XII(cen15) XVI:IX:III:I:VI:II (cen3) XI:VII:XIII:XIV(cen7)*	3’
14	yJL402	Chr15R was fused to chr04L, with cen4 deleted with parental strain yJL381	*MATa his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 XVI:IX:III:I:VI:II:XI:VII:XIII:XIV (cen7) V:VIII:X:XV:IV:XII(cen15)*	2

Extended Data Table 3 | Variants identified from genome sequencing of different *n* strains

N	Chromosome	Position	Reference	Mutation	ORF (amino acids mutation)
12	*Chr II*	8972	G	A	Non-coding
		8975	T	C	Non-coding
		557493	A	C	*AMN1* (YBR158W, L317F)
	Chr XIV	577042	ATCGGTGGTTAAACAA	A	Non-coding
8	*Chr II*	8972	G	A	Non-coding
		8975	T	C	Non-coding
		322190	G	A	*QDR3*(YBR043C, P586S)
		557493	A	C	*AMN1* (YBR158W, L317F)
	Chr XIV	577042	ATCGGTGGTTAAACAA	A	Non-coding
4	*Chr II*	8972	G	A	Non-coding
		8975	T	C	Non-coding
		9022	GAAA	G	Non-coding (poly A region)
		322190	G	A	*QDR3*(YBR043C, P586S)
		557493	A	C	*AMN1* (YBR158W, L317F)
	Chr IV	449785	A	C	Non-coding
		979062	C	T	*EXG2*(YDR261C, A49T)
	Mitochondria	2483	CTAATAAT	C	Non-coding (Tandem repeat region)
		3916	C	CT	Non-coding (poly T region)
		54569	A	AG	Non-coding (poly G region)
	Chr VII	614706	A	C	*ADE6*(YGR061C, C422G)
	Chr X	648070	C	A	*ATP2*(YJR121W, A158E)
	Chr XIV	8046	CCGTGA	C	Non-coding
		8052	T	TCTCACA	Non-coding
		246906	C	T	*RRG9*(YNL213C, R67K)
		577042	ATCGGTGGTTAAACAA	A	Non-coding
	Chr XV	976987	AT	A	Non-coding (poly T region)
2	*Chr II*	8972	G	A	Non-coding
		8975	T	C	Non-coding
		322190	G	A	*QDR3*(YBR043C, P586S)
		557493	A	C	*AMN1* (YBR158W, L317F)
	Chr III	114916	AT	A	Non-coding
	Chr IV	979062	C	T	*EXG2*(YDR261C, A49T)
	Mitochondria	2483	CTAATAAT	C	Non-coding (Tandem repeat region)
		3916	C	CT	Non-coding (poly T region)
		29767	T	A	Non-coding (in a AT rich region)
		54569	A	AG	Non-coding (poly G region)
	Chr V	193248	G	T	*ISC1* (YER019W, synonymous mutation)
	Chr VII	614706	A	C	*ADE6*(YGR061C, C422G)
	Chr VIII	133864	A	T	Non-coding
	Chr X	648070	C	A	*ATP2*(YJR121W, A158E)
	Chr XIV	8046	CCGTGA	C	Non-coding
		8052	T	TCTCACA	Non-coding
		246906	C	T	*RRG9*(YNL213C, R67K)
		577042	ATCGGTGGTTAAACAA	A	Non-coding
	Chr XV	976987	AT	A	Non-coding (poly T region)

Mutations inherited from previous steps are shaded.

Extended Data Table 4 | Gene expression changes in $n = 2$, 4 and 8 strains

a

	N=2	N=4	N=8
Expression decreases	RMD6*, YEL077C*, YLR460C*, AIF1*, DLD3*, COS10*	GEX2*, YRF1-4*, DLD3*, YLR460C*, AIF1*, YEL073C*, YEL077C*, RMD6*, COS7*	YLR460C*, YDR541C*, COS7*
Expression increases	YKL187C, SPG4, AAD10*, RPS9A, YJL045W, YLR281C, COS8*, OYE3, HXT5, IMD2*, HSP12, HMX1, HBT1, PHO11*, SPG1, GRE1, SIP18, YIR042C*, PIR3, FMP45, GND2, RNR3, YDR034W-B, HXT9*, PHM7, PUT4, COS1*, CDA1, TKL2, HUG1, COS4*, VBA3*	YER188W*, COS8*, IMD2*, YIR042C*, COS1*, COS4*, VBA3*	YIR042C*, COS1*, COS8* YER188W*, IMD2*, COS4*

b

Gene Systematic name	N=2	N=4	N=8
COS4/YFL062W	significant	significant	significant
YFR057W/YFR057W	significant	significant	significant
YPS5/YGL259W	significant	significant	significant
IMD2/YHR216W	significant	significant	significant
YGL258W-A /YGL258W-A	significant	not significant	significant
GEX1/YCL073C	significant	significant	not significant
VBA3/YCL069W	significant	significant	not significant
PAU4/YLR461W	significant	not significant	not significant
FDH1/YOR388C	significant	not significant *	not significant *
THI5/YFL058W	not significant	not significant	not significant
AAD15/YOL165C	Removed by design	Removed by design	not significant *
Fisher's Exact test	p-value= 9.34 x e-12	p-value= 1.16 x e-6	p-value= 8.66 x e-7

c

	Significantly changed in n=2 strain	Significantly changed in n=4 strain	Significantly changed in n=8 strain	All genes
Subtelomeric genes	15	16	9	325
Non-subtelomeric genes	24	0	0	6360
P value (Fisher's Exact test)	2.212 x e-10	2.20 x e-16	1.729 x e-12	

a, Under- or overexpressed genes in $n = 2$, 4 and 8 strains. Genes in the fused chromosome arm are labelled as red. Genes in remaining unfused chromosome arms are in black. *Genes located near a telomere in $n = 16$ strain (within 20 kb). **b**, Comparison of the expression changes of 21 subtelomeric genes, with changes in Sir2Δ, Sir3Δ and Sir4Δ strains in ref. [40]. Twenty-one subtelomeric genes that are significantly expressed in ref. [40] are listed in the first column. Whether they are significantly expressed in our strains ($n = 2$, $n = 4$ and $n = 8$) or not is listed in the second, third and fourth columns. The last row shows the correlation between fusion chromosome strains and Sir protein depletion strains. *IMD1, YAR075W, YCL076W, YCL075W, YCL074W, YFL063W, YNL337W* are removed by design. *YCL068C, YCL065W* and *YCL064C* are part of the silent *HML* locus. *Unfused telomeres. The *P* value was calculated by a one-sided Fisher's exact test. **c**, Transcriptome change enrichments in sub-telomeres are significant in $n = 2$, 4 and 8 cells, calculated by one-sided Fisher's exact test.

Extended Data Table 5 | Primers for PCR verification of fusion chromosomes

Fusion step	Purpose	Forward primer (5'-3')	Reverse primer (5'-3')
1	*chr III-I* junction	TGAAACGCACTGCAACATCTG	GGCTCAATCATCTACCGCATGG
	CEN 1 deletion	CAACCAAACGTCCTCTTCTCTC	ACGATACATGGACTGACTCAAG
2	*chr IX-III* junction	TTAAGGTGCGACCGGCAATG	CCTGTTTCGGCACTTGAGTC
	CEN 9 deletion	CAACGAATTTCTCTCCGCCAGG	CACTTCAACAGTGCCAAAGACTCTAC
3	*chr V-VIII* junction	TACGCCAAGTCGGTCAGGTC	TCCGAACTTGGTGTGTCTTCAG
	CEN 5 deletion	ACCTCCTAGCACTTCGTAATG	GCTATTTATGTGCGGCTTTGTC
4	*chr VI-II* junction	CAGATCCTTTCGCATTCCTACTTG	TCGTCGATGGTATTGGTGTAGAG
	CEN 6 deletion	TTGGGCGATGGAAGAGGTAAAG	ACTTTCAACGCAAGAGCAAGAC
5	*chr XIII-XIV* junction	CCAAGACTCTCACCTGCGAC	GCAATGGCTCAGTAACCTCG
	CEN 14 deletion	CTGATGGACTCCGTAGAGAGC	AGGGTAGCATAAACCTGCTG
6	*chr XVI-IX* junction	ACATTTGGGCCGTTGCTAGAAG	TATAGTCGGGCCTAGTTGCACTC
	CEN 16 deletion	GGTTGAAGGAGTTAGTTTGTCG	GCCGCTTTGATGATTCTGCTTTAG
7	*chr VIII-X* junction	GAACCGCTTCTGCTCAACTAG	CAATGACGGTGTTCGTGAAGC
	CEN 10 deletion	CTCAGAAGGGAATTTCGTAAGC	CCAGTTTAGTTGTTGTGGATGC
8	*chr I-VI* junction	CGTCAGCAGCGTCAGTAACTC	GCTGCAACAACTTCCCAATCATG
	CEN 2 deletion	GGACTGAAAGCCAGTAACAAGC	TTCTCGTACCAAGCCGGTTC
9	*chr VII-XIII* junction	AAAGTTTCCACCAGACGCTAAG	CTACACTCGAACTCTGTTTCTCTC
	CEN 13 deletion	AGGCTTTCGATTACCATGTGC	CTAAGGTAGCCAGAACTTCTCATC
10	*chr IV-XII* junction	GCGTCACTTCTAAGAACAAGACTGC	ATGGTGAGAGATGGGTGATGGAG
	CEN 12 deletion	GACAACCAAACTGGTGTATGC	TGCCATCATCTACTTCCTTTCC
11	chr XI-VII junction	GCGAAAGCGAAGCCAATGTG	GCCATTAGCCTTCTATGTGTTC
	CEN 11 deletion	GAACGACATTAACGGATACGCAAC	TGAAGAAGGTCAACATGAGGATGG
12	*chr X-XV* junction	AATGCTGTGACACGCAGATAC	GGTACGCTCACCTCGTAAGTC
	CEN 8 deletion	ACCCTCAGGTTGCTATGACG	ACGCACGAGCGAATTAACATTCC
13*	*chr II-XI* junction	GCTGCAACAACTTCCCAATCATG	TTTGCCAACACGAAAGGAACTC
14	*chr XV-IV* junction	GGTAGTAACCAACTCGTATCCCTG	TGGCATTCCTCTTTCACTTTCGTC
	CEN 4 deletion	AGTGGTTGACATGCTGGCTAG	GGCCTCAAGAAAGAAACCTCTATG

*Centromere 3 was inactivated by the deletion of only 3 bp CGG. Sanger sequencing was used to verify the deletion of CGG rather than PCR amplification.

LETTER

https://doi.org/10.1038/s41586-018-0407-5

Gamete fusion triggers bipartite transcription factor assembly to block re-fertilization

Aleksandar Vještica[1], Laura Merlini[1], Pedro Junior Nkosi[1] & Sophie G. Martin[1]*

The ploidy cycle, which is integral to sexual reproduction, requires meiosis to halve chromosome numbers as well as mechanisms that ensure zygotes are formed by exactly two partners[1–4]. During sexual reproduction of the fungal model organism *Schizosaccharomyces pombe*, haploid P and M cells fuse to form a diploid zygote that immediately enters meiosis[5]. Here we reveal that rapid post-fusion reconstitution of a bipartite transcription factor blocks re-fertilization. We first identify mutants that undergo transient cell fusion involving cytosol exchange but not karyogamy, and show that this drives distinct cell fates in the two gametes. The P partner undergoes lethal haploid meiosis, whereas the M cell persists in mating. The zygotic transcription that drives meiosis is rapidly initiated first from the P parental genome, even in wild-type cells. This asymmetric gene expression depends on a bipartite complex formed post-fusion between the cytosolic M-cell-specific peptide Mi and the nuclear P-cell-specific homeobox protein Pi[6,7], which captures Mi in the P nucleus. Zygotic transcription is thus poised to initiate in the P nucleus as fast as Mi reaches it after fusion, a design that we reconstruct using two synthetic interactors localized to the nucleus and the cytosol of two respective partner cells. Notably, delaying zygotic transcription—by postponing Mi expression or deleting its transcriptional target in the P genome—leads to zygotes fusing with additional gametes, thus forming polyploids and eventually aneuploid progeny. The signalling cascade to block re-fertilization shares components with, but bifurcates from, meiotic induction[8–10]. Thus, a cytoplasmic connection upon gamete fusion leads to asymmetric reconstitution of a bipartite transcription factor to rapidly block re-fertilization and induce meiosis, ensuring genome maintenance during sexual reproduction.

While studying fusion-defective mutants[11], we discovered an interesting post-fusion asymmetry between mating partners. Mating mixtures of cells that lack the p21-activated kinase Pak2 (also known as Shk2) showed many unfused partners and a prolonged actin fusion focus lifetime[12] (Fig. 1a, Extended Data Fig. 1a). *pak2*Δ matings also produced about 10% aberrant asci, apparently either derived from three cells and containing more than four spores (Fig. 1a, type IIIa), or derived from a single cell that underwent sporulation (Fig. 1a, type IIIb). Three lines of evidence revealed that these asci are the result of meiosis and sporulation taking place in haploid cells.

First, live-imaging showed that aberrant asci were formed upon mating of partners that had not undergone karyogamy. The spore-forming partner displayed a long microtubule bundle—characteristic of meiotic prophase[13,14]—followed by two rounds of spindle formation, indicative of meiosis I and meiosis II (Fig. 1b, Extended Data Fig. 1b, Supplementary Video 1a). The other cell maintained interphase microtubules and the ability to mate with another partner, thus yielding type IIIa asci. Second, fluorescent labelling of a chromosomal locus revealed that spore-forming cells contained only two loci distributed between up to four spores, which indicates aneuploidy (Extended Data Fig. 1c). By contrast, every wild-type spore contained one fluorescent locus (Extended Data Fig. 1c; see also ref. [15]). Third, flow cytometry analysis of DNA-stained spores showed *pak2*Δ spores with sub-1C DNA content (Fig. 1c). Thus, aberrant *pak2*Δ asci form upon meiosis and sporulation in a single, haploid cell.

Haploid meiosis was provoked by transient cell fusion: *pak2*Δ cells that produced aberrant asci initially fused and exchanged cytosolic GFP, but their fusion pore re-sealed, as indicated by persistent unequal levels of GFP between partners ($n > 20$, Fig. 1d, Extended Data Fig. 1d, e, Supplementary Video 1b, c). Transient fusion was required to induce haploid meiosis because fully blocking fusion in formin *fus1*[12,16] mutants prevented aberrant asci formation (Extended Data Fig. 1f). Pak2 did not regulate meiosis directly, as *pak2*Δ diploids sporulated normally (Extended Data Fig. 1g). In crosses between wild-type and *fus1*Δ cells—which exhibit partial fusion defects[12,16]—we also observed transient fusions followed by haploid meiosis in one partner, whether this was *fus1*Δ or wild type (Extended Data Fig. 1f, h, i, Supplementary Video 1d). We conclude that, independent of genotype, transient cell fusion induces meiosis in one partner cell. This is consistent with observations that the merging of M and P cytosols induces meiosis independently of ploidy or karyogamy[17–19].

Transient fusion induced haploid meiosis strictly in the P cell (in 64 out of 64 cases for *pak2*Δ, and in 9 out of 9 cases between *fus1*Δ and wild type). In transiently fusing pairs, the master meiotic regulator Mei3[10,20] was induced only in P cells, as assessed by tagging *mei3* with a fast-folding GFP variant (sfGFP)[21] (Fig. 2a, Supplementary Video 5). M genome *mei3*–sfGFP produced fluorescence only upon complete fusion (Fig. 2b, Supplementary Video 2a). Importantly, *mei3* was also asymmetrically expressed in wild-type zygotes: Mei3–sfGFP expression was significantly more rapid when encoded in the P genome than in the M genome (Fig. 2c, Extended Data Fig. 2a, Supplementary Video 2b). Thus, zygotes first express the meiotic inducer Mei3 from the P genome.

Mei3 expression is under the regulation of two cell-type-specific factors, the P-specific homeodomain protein Pi and the M-specific 42-amino-acid peptide Mi[6,7]. We found that Pi and Mi formed complexes insensitive to nuclease treatment both in vivo and when expressed in bacteria (Fig. 2d, e). Mi co-expression also stabilized Pi levels (Fig. 2d). Chromatin immunoprecipitation experiments further showed that Pi and Mi associate with the *mei3* promoter, but only when both proteins are present (Fig. 2f, Extended Data Fig. 2c). Thus, the Pi–Mi complex—but not its individual subunits—binds chromatin and induces *mei3* expression.

To test whether Pi and Mi govern the asymmetric *mei3* induction from parental genomes, we swapped their coding sequences. This swap produced a phenotype reversal, with Mei3 expressed first from the M genome (Fig. 2g, Extended Data Fig. 2b, Supplementary Video 3a) and transient fusion inducing haploid meiosis in M cells (in 13 out of 14 transient fusions, Extended Data Fig. 2d, e, Supplementary Video 3b). The swap did not affect mating and sporulation efficiencies nor spore viability (Extended Data Fig. 2f–h). Thus, asymmetric zygotic *mei3* transcription is governed by Mi and Pi, with the Pi-expressing genome inducing *mei3* first.

We visualized Pi and Mi tagged with sfGFP at their N- and C-termini, respectively, expressed from their native genomic loci. Mi–sfGFP

[1]Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland. *e-mail: Sophie.Martin@unil.ch

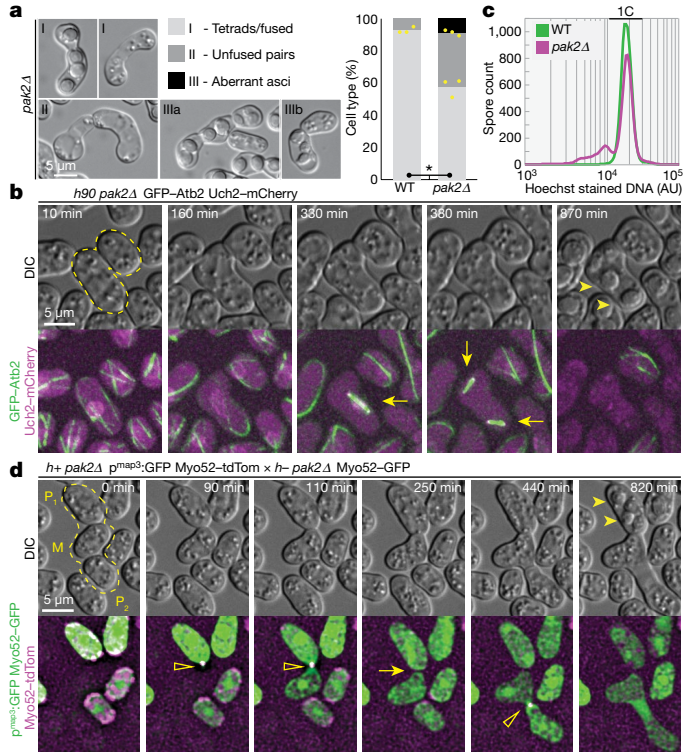


Fig. 1 | Transient cell fusion induces haploid meiosis and aneuploidy. **a**, Quantification (right) of shown phenotypes (left) in homothallic wild-type and *pak2*Δ matings. $*P = 4.81 \times 10^{-3}$ ($n = 3$ experiments, with $n \approx 500$ cell pairs each), Welch's test. WT, wild type. **b**, Homothallic *pak2*Δ mating cells expressing GFP–α-tubulin (green) and nuclear Uch2–mCherry (magenta). Note the lack of karyogamy in the outlined *pak2*Δ pair though one partner forms meiotic spindles (arrows) and spores (arrowhead), while the other maintains interphase microtubules. Extended Data Fig. 1b shows wild type. **c**, Flow cytometry analysis of DNA in spores produced by wild-type and *pak2*Δ matings ($n = 10{,}000$). AU, arbitrary units. **d**, Transient fusion in *pak2*Δ matings observed as P-cell-expressed GFP enters into the M cell, followed by fusion pore sealing (arrow) and spore formation (filled arrowheads) in the P_1 cell. The fusion focus (empty arrowhead) is labelled by M cell Myo52–GFP and P cell Myo52–tdTomato. Extended Data Fig. 1d shows wild type.

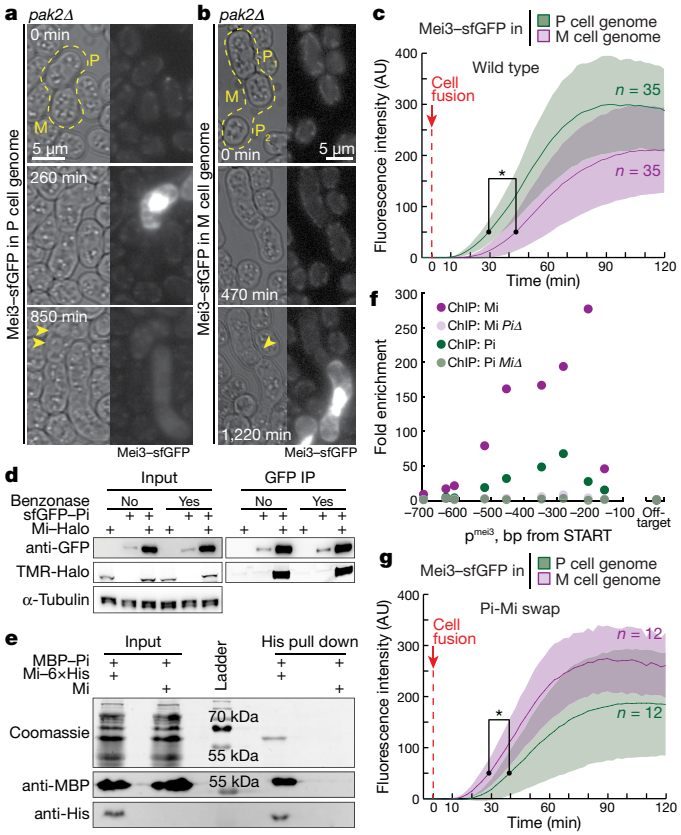


Fig. 2 | The meiotic inducer Mei3 is rapidly induced from the genome of cells expressing homeodomain factor Pi. **a**, GFP-tagging of P genome *mei3* leads to Mei3–sfGFP expression in the *pak2*Δ spore-forming P cell (arrowhead), but not in the M cell. $n > 10$ cells. **b**, GFP-tagging of M genome *mei3* shows no expression during transient fusion of *pak2*Δ M and P_1 cells—which produces spores (arrowhead)—but strong expression when M and P_2 cells completely fuse. $n > 10$ cells. **c**, Mei3–sfGFP signal from zygotes of heterothallic wild-type cells with *mei3* tagged in only one parental genome. Time zero is cell fusion, visualized by transfer of cytosolic mCherry. Solid lines represent average; shaded areas include two standard deviations. $*P = 1.39 \times 10^{-8}$, Kruskal–Wallis test. **d**, Western blot of cell lysates (input) from interphase cells overexpressing sfGFP–Pi, Mi–Halo or both, immunoprecipitated with GFP affinity beads (GFP IP) and treated or not with benzonase. Tubulin served as loading control. Note Mi co-purification with Pi and Pi stabilization by Mi co-expression. **e**, Coomassie-stained SDS–PAGE and western blots of benzonase-treated lysates (input) obtained from *Escherichia coli* co-expressing MBP–Pi and either Mi or Mi–6×His, purified with cobalt resin (His pull down). **f**, Chromatin immunoprecipitation (ChIP) for endogenous sfGFP-tagged Pi and Mi, in the presence or absence of the binding partner, analysing *mei3* promoter and off-target region association. Extended Data Fig. 2c shows replicates. **g**, Mei3 induction upon mating of cells with swapped Mi–Pi. Data processed as in **c**. $*P = 1.8 \times 10^{-3}$, Kruskal–Wallis test. Gels are available in Supplementary Fig. 1.

produced a faint cytoplasmic signal in early mating M cells. Notably, fusion induced rapid Mi–sfGFP accumulation in the P nucleus followed by a delayed M nucleus entry (Fig. 3a, Extended Data Fig. 3a, Supplementary Video 4a, b). In transiently fusing *pak2*Δ cells, Mi–sfGFP accumulated in the nucleus of the P cell—which underwent haploid sporulation (Fig. 3b, Supplementary Video 4c)—but not in the M nucleus. Thus, Mi nuclear accumulation requires a P-cell-specific factor that remains asymmetrically distributed upon transient fusion. Three lines of evidence showed that this P-cell-specific factor is Pi. First, Pi was nuclear: sfGFP–Pi produced a very low-intensity signal in both P cell and zygote nuclei (Extended Data Fig. 3b, Supplementary Video 5a). Pi nuclear localization was more evident in fusion-defective *fus1*Δ cells (Extended Data Fig. 3c, Supplementary Video 5a) or upon overexpression in interphase cells (Fig. 3c). Second, cytosolic Mi–mCherry that was overexpressed in interphase cells accumulated in the nucleus upon sfGFP–Pi co-overexpression (Fig. 3c). Mi–mCherry expression also led to increased sfGFP–Pi levels (1.49 ± 0.42-fold, $n = 20$, $P = 7.95 \times 10^{-7}$; Kruskal–Wallis test), as observed above (see Fig. 2d). Finally, deletion of Pi prevented Mi–sfGFP nuclear enrichment after fusion (Fig. 3d, Supplementary Video 5b).

We investigated the causes of Mi asymmetric nuclear localization. The asymmetry reversal associated with the Mi–Pi swap indicates that asymmetry is independent of other cell-type-specific factors (Fig. 2g, Extended Data Fig. 2d, e). The 5-kDa Mi peptide may diffuse more rapidly than its 19-kDa partner Pi. However, tagging Mi with 27-kDa sfGFP did not abrogate asymmetry, which suggests that asymmetry is not simply reliant on protein size. We thus investigated the possible contribution of Pi and Mi localization to the nucleus and cytosol. We mimicked these differentially localized interactors with the synthetic hetero-specific short coiled-coils SynZip3 and SynZip4[22], the former fused to mCherry, the latter to sfGFP and a nuclear localization signal, each expressed in a separate partner. Upon cell fusion, the cytosolic marker accumulated in the partner's nucleus immediately, whereas equilibration of the nuclear marker was delayed—mimicking Mi and Pi behaviours (Fig. 3e, Supplementary Video 6). Such asymmetry was even more notable upon transient fusion (Extended Data Fig. 3d, Supplementary Video 6). No asymmetric localization was observed in the absence of a nuclear localization signal (Extended Data Fig. 3e, Supplementary Video 6). Thus, asymmetry is inherent to the distinct

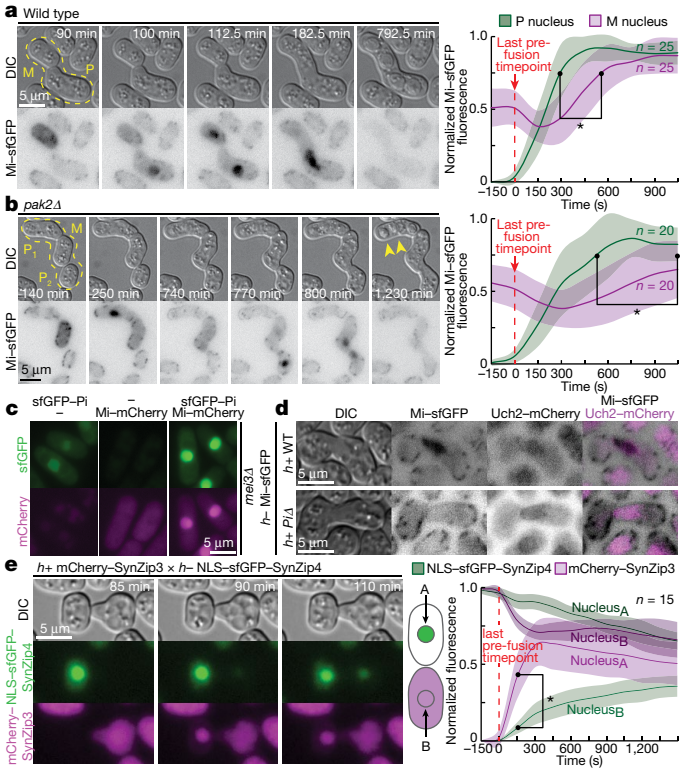


Fig. 3 | Meiotic regulators Pi and Mi localize asymmetrically in early zygotes. a, **b**, M-cell-specific cytosolic Mi–sfGFP rapidly accumulates in the P nucleus after partners fuse, and only subsequently in the M nucleus in wild type (**a**) and *pak2*Δ (**b**). Mi–sfGFP accumulates in *pak2*Δ P_1 nucleus but not in M nucleus upon transient fusion. The punctate cortical signal is background—probably mitochondrial—fluorescence. Right panels quantify average (lines) and two standard deviations (shaded area) of nuclear Mi–sfGFP. In **a**, $*P = 1.79 \times 10^{-7}$, Kruskal–Wallis test. In **b**, $*P < 8.97 \times 10^{-5}$, Kruskal–Wallis test. **c**, Exponentially growing *mei3*Δ cells overexpressing sfGFP–Pi, Mi–mCherry or both proteins. Note nuclear Mi–mCherry upon Pi co-expression and higher sfGFP–Pi signal upon Mi co-expression. **d**, Zygotes produced by Mi–sfGFP-expressing M cells and wild-type (top) or *Pi*Δ (bottom) P cells. Uch2–mCherry labels nuclei. **e**, Mating of wild-type nuclear localization signal (NLS)–sfGFP–SynZip4-expressing and mCherry–SynZip3-expressing cells. Upon fusion, the cytosolic red fluorophore rapidly accumulates in the partner's nucleus, whereas nuclear green fluorescence exchange is delayed. Some vacuolar red signal does not readily exchange. Right panel quantifies nuclear signals, and is presented as in **a**. $*P = 3.39 \times 10^{-6}$, Kruskal–Wallis test.

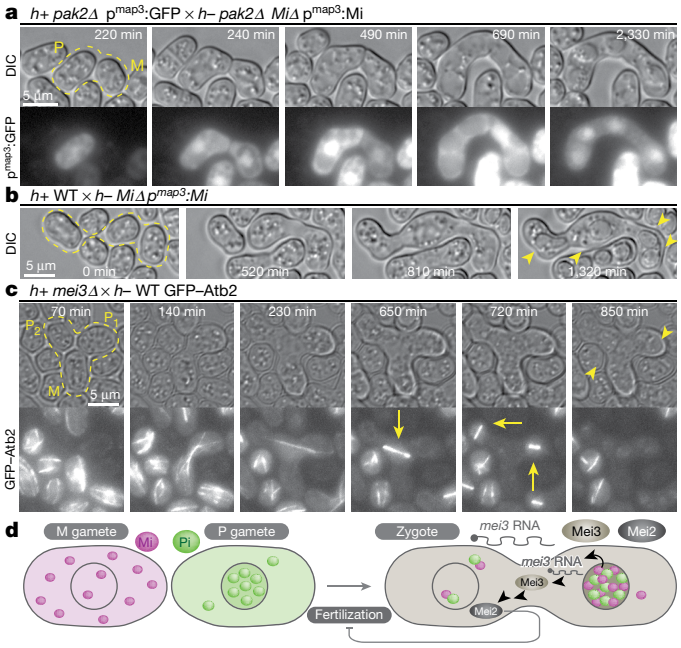


Fig. 4 | Rapid induction of Mei3–Mei2 signalling suppresses mating in zygotes and prevents polyploid formation. a, Mating of *pak2*Δ M cells expressing Mi from P cell specific p^{map3} promoter with cytosolic GFP-labelled P cells. Transient fusion, observed as GFP exchange and signal build-up in only one partner, never induced meiosis ($n = 10$). The cells shown persistently mate until complete fusion, followed by sporulation. **b**, Zygote fusing with an additional partner upon mating of wild-type P cells with M cells expressing Mi under p^{map3} promoter. Note formation of spores in the now-triploid zygote. **c**, Mating of *mei3*Δ P cells and GFP–α-tubulin-expressing M cells. Note fusion of three partners, formation of meiotic spindles (arrows) and spores (arrowheads). **d**, Model for the asymmetric localization of the bipartite Pi–Mi transcriptional activator in early zygotes, which results in fast P genome *mei3* expression followed by Mei2 activation to block re-fertilization.

sub-cellular localization of Mi and Pi, a design that promotes rapid *mei3* transcription from the P genome.

In summary, cell fusion reconstitutes a bipartite Pi–Mi transcription factor. Pi in the P nucleus rapidly traps cytosolic Mi (Fig. 4d). The Pi–Mi complex then binds the *mei3* promoter and induces Mei3 expression from the P genome. Mei3 expression from the M-genome is postponed owing to the delayed exchange of Pi between partners.

To address the role of rapid Mei3 induction, we delayed Mei3 expression in two ways. We first built a transcriptional delay by placing Mi in M cells under control of a P-specific promoter, such that Mi is expressed only after successful fusion. This delayed Mei3–sfGFP expression by about 30 min (Extended Data Fig. 4a, Supplementary Video 7a). Second, we simply deleted P genome *mei3*, such that Mei3 is expressed only from the M genome and thus with an approximately 15-min delay (Fig. 2c). Though neither approach affected mating or sporulation efficiencies upon fusion of otherwise wild-type cells (Extended Data Fig. 4b), both prevented haploid meiosis upon transient fusion of *pak2* mutants. Instead, *pak2*Δ partners persistently attempted to form a zygote (Fig. 4a, Extended Data Fig. 4c, Supplementary Video 7b). As expected, deleting *mei3* in *pak2*Δ M cells did not affect haploid meiosis in the P cell (Extended Data Fig. 4d, Supplementary Video 7b).

Remarkably, delays in Mei3 expression in otherwise wild-type crosses led some zygotes to engage additional partners (Fig. 4b, c, Extended Data Fig. 4e–g, Supplementary Video 8) ($n = 18$ out of about 7,000 events for both delay in Mi expression ($n = 4$ experiments) and *mei3* deletion in P cells ($n = 3$ experiments), with P values of 0.033 and 0.031, respectively, compared to wild type; Welch's test). Multi-partner zygotes entered meiosis—as shown by microtubule labelling—and formed four, probably aneuploid spores (Fig. 4b, c, Supplementary Video 8). Polyploid zygotes were never observed in wild type ($n > 10{,}000$ events) and very rarely when *mei3* was deleted from the M genome (2 of $n > 7{,}000$ events ($n = 3$); P value = 0.42 compared to wild type; Welch's test). Thus, the rapid onset of Mei3 expression from the zygotic P genome prevents the formation of polyploid zygotes, and consequently the formation of aneuploid spores.

A larger fraction of zygotes (about 1%) fused with additional partners upon *mei3*, *Pi* or *Mi* deletion in homothallic (self-fertile) strains, and these predictably did not enter meiosis (Extended Data Fig. 5a–d, Supplementary Video 9). In addition, >30% of zygotes exhibited growth that was suggestive of mating behaviour (Extended Data Fig. 5b, Supplementary Video 9). Mei3 alleviates repression of the RNA-binding protein Mei2[9], which—in complex with *meiRNA* RNA encoded by the *sme2* locus[8]—induces meiosis by promoting the expression of the meiotic forkhead transcription factor Mei4[23,24]. Deletion of *mei2* or mutation of its C-terminal RNA-recognition motif[8] resulted in about 10% multi-partner zygotes (Extended Data Fig. 5e, f, Supplementary Video 9; see also ref. [11]). Approximately 90% of *mei2*Δ*mei3*Δ double mutant zygotes grew mating projections, and about 16% fused with a third partner (Extended Data Fig. 5g, Supplementary Video 9). By contrast, *mei4*Δ and *sme2*Δ zygotes never engaged additional partners, though both arrested before spore formation (Extended Data Fig. 5h, i).

We conclude that the re-fusion block depends on Mei2 signalling—probably through RNA-binding, but not with meiRNA. Thus, mating repression in zygotes shares components with, but bifurcates from, meiotic induction.

Mechanisms preventing re-fertilization ensure ploidy maintenance across generations in evolutionarily divergent phyla. These mechanisms generally involve rapid changes, such as the release of cortical granules[4] and shedding of surface receptors in mammals[1], membrane depolarization in amphibians[4,25] and the degradation of pollen guidance cues in plants[26]. Our results provide a first glimpse into fungal re-fertilization blocks, by showing that yeast zygotes rapidly initiate transcription to discontinue mating. Whereas Mi peptides are fast-evolving peptides that we could only identify in the *Schizosaccharomyces* lineage by genome position, Pi-like homeodomain proteins are ancestral eukaryotic proteins that are extensively used in developmental processes[27]. Their primordial role may be in the haploid-to-diploid transition[28], which suggests that their rapid post-fusion activation is used elsewhere to block re-fertilization. For instance, rapid post-fusion homeodomain complex formation and nuclear translocation has previously been observed in *Chlamydomonas reinhardtii*[29].

The bipartite design of the Mi–Pi transcription factor—which is reminiscent of hormone nuclear receptors and their ligands[30]—favours post-fusion activation speed, limited only by the rate at which Mi reaches the nuclear-localized homeodomain Pi protein. Because this simple two-component system inherently leads to asymmetric zygotic expression, asymmetry may follow from the selective pressure to rapidly block re-fertilization. As even a transient cytoplasmic connection is sufficient for formation of an active complex, such transcription factor design may also be used to impart cell fate changes upon—or build sensors to monitor—other instances of transient cytoplasmic bridge formation.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0407-5.

Received: 24 March 2017; Accepted: 25 June 2018;
Published online 8 August 2018.

1. Bianchi, E., Doe, B., Goulding, D. & Wright, G. J. Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487 (2014).
2. Völz, R., Heydlauff, J., Ripper, D., von Lyncker, L. & Groß-Hardt, R. Ethylene signaling is required for synergid degeneration and the establishment of a pollen tube block. *Dev. Cell* **25**, 310–316 (2013).
3. Bleckmann, A., Alter, S. & Dresselhaus, T. The beginning of a seed: regulatory mechanisms of double fertilization. *Front. Plant Sci.* **5**, 452 (2014).
4. Cheeseman, L. P., Boulanger, J., Bond, L. M. & Schuh, M. Two pathways regulate cortical granule translocation to prevent polyspermy in mouse oocytes. *Nat. Commun.* **7**, 13726 (2016).
5. Merlini, L., Dudin, O. & Martin, S. G. S. Mate and fuse: how yeast cells do it. *Open Biol.* **3**, 130008 (2013).
6. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. Four mating-type genes control sexual differentiation in the fission yeast. *EMBO J.* **7**, 1537–1547 (1988).
7. Willer, M. et al. Two-step activation of meiosis by the mat1 locus in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **15**, 4964–4970 (1995).
8. Watanabe, Y. & Yamamoto, M. *S. pombe mei2*$^{+}$ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell* **78**, 487–498 (1994).
9. Watanabe, Y., Shinozaki-Yabana, S., Chikashige, Y., Hiraoka, Y. & Yamamoto, M. Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast. *Nature* **386**, 187–190 (1997).
10. Li, P. & McLeod, M. Molecular mimicry in development: identification of *ste11*$^{+}$ as a substrate and *mei3*$^{+}$ as a pseudosubstrate inhibitor of *ran1*$^{+}$ kinase. *Cell* **87**, 869–880 (1996).
11. Dudin, O. et al. A systematic screen for morphological abnormalities during fission yeast sexual reproduction identifies a mechanism of actin aster formation for cell fusion. *PLoS Genet.* **13**, e1006721 (2017).
12. Dudin, O. et al. A formin-nucleated actin aster concentrates cell wall hydrolases for cell fusion in fission yeast. *J. Cell Biol.* **208**, 897–911 (2015).
13. Ding, D. A rush hour towards sexual reproduction: the chromosome dynamics during meiosis. *Chin. Sci. Bull.* **56**, 3500–3503 (2011).
14. Ding, D. Q., Yamamoto, A., Haraguchi, T. & Hiraoka, Y. Dynamics of homologous chromosome pairing during meiotic prophase in fission yeast. *Dev. Cell* **6**, 329–341 (2004).
15. Yamamoto, T. G., Chikashige, Y., Ozoe, F., Kawamukai, M. & Hiraoka, Y. Activation of the pheromone-responsive MAP kinase drives haploid cells to undergo ectopic meiosis with normal telomere clustering and sister chromatid segregation in fission yeast. *J. Cell Sci.* **117**, 3875–3886 (2004).
16. Petersen, J., Weilguny, D., Egel, R. & Nielsen, O. Characterization of fus1 of *Schizosaccharomyces pombe*: a developmentally controlled function needed for conjugation. *Mol. Cell. Biol.* **15**, 3697–3707 (1995).
17. Polakova, S., Benko, Z., Zhang, L. & Gregan, J. Mal3, the *Schizosaccharomyces pombe* homolog of EB1, is required for karyogamy and for promoting oscillatory nuclear movement during meiosis. *Cell Cycle* **13**, 72–77 (2014).
18. Yamashita, A., Fujita, Y. & Yamamoto, M. Proper microtubule structure is vital for timely progression through meiosis in fission yeast. *PLoS ONE* **8**, e65082 (2013).
19. Gutz, H. "Twin meiosis" and other ambivalences in the life cycle of *Schizosaccharomyces pombe*. *Science* **158**, 796–798 (1967).
20. McLeod, M. & Beach, D. A specific inhibitor of the *ran1*$^{+}$ protein kinase regulates entry into meiosis in *Schizosaccharomyces pombe*. *Nature* **332**, 509–514 (1988).
21. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
22. Reinke, A. W., Grant, R. A. & Keating, A. E. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* **132**, 6025–6031 (2010).
23. Murakami-Tonami, Y. et al. Mei4p coordinates the onset of meiosis I by regulating *cdc25*$^{+}$ in fission yeast. *Proc. Natl Acad. Sci. USA* **104**, 14688–14693 (2007).
24. Mata, J., Wilbrey, A. & Bähler, J. Transcriptional regulatory network for sexual differentiation in fission yeast. *Genome Biol.* **8**, R217 (2007).
25. Jaffe, L. A. Fast block to polyspermy in sea urchin eggs is electrically mediated. *Nature* **261**, 68–71 (1976).
26. Beale, K. M., Leydon, A. R. & Johnson, M. A. Gamete fusion is required to block multiple pollen tubes from entering an *Arabidopsis* ovule. *Curr. Biol.* **22**, 1090–1094 (2012).
27. Derelle, R., Lopez, P., Le Guyader, H. & Manuel, M. Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. *Evol. Dev.* **9**, 212–219 (2007).
28. Bowman, J. L., Sakakibara, K., Furumizu, C. & Dierschke, T. Evolution in the cycles of life. *Annu. Rev. Genet.* **50**, 133–154 (2016).
29. Lee, J.-H., Lin, H., Joo, S. & Goodenough, U. Early sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. *Cell* **133**, 829–840 (2008).
30. Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors, RXR, and the big bang. *Cell* **157**, 255–266 (2014).

Acknowledgements We thank M. Bühler, T. Kuzdere, F. Bendezú, S. Pelet, M. Marek, B. Arcangioli and G. Thon for strains and experimental advice, and R. Benton, T. Andersen, S. Gruber, S. Mitri, J.-W. Veening and members of the Martin laboratory for manuscript suggestions. An EMBO long-term fellowship to A.V., ERC Consolidator Grant (CellFusion) and Swiss National Science foundation grant (31003A_155944) to S.G.M. supported this work.

Reviewer information *Nature* thanks S. Grewal, J. Heitman and O. Nielsen for their contribution to the peer review of this work.

Author contributions S.G.M. and A.V. conceived the project, designed experiments and wrote the manuscript. A.V. performed experiments, with P.J.N. assisting with pull-downs and L.M. with constructing strains.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0407-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0407-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to S.G.M.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Growth conditions. The growth conditions used for experiments are detailed in ref. [31] and the overview of experimental procedure is presented in Extended Data Fig. 6a. In brief, freshly streaked cells were inoculated into MSL + N medium[32] and incubated overnight at 25 °C with 200 rpm shaking to exponential phase. The following evening cultures were diluted to $OD_{600} = 0.025$ in 20 ml of MSL + N medium and incubated overnight at 30 °C with 200 rpm agitation to exponential phase. Experiments on exponentially growing cultures were performed at this point. For time-lapse imaging of mating, homothallic cells or 1:1 mixtures of heterothallic cells were pelleted for one minute at 1000*g* in microcentrifuge tubes and washed three times in 1 ml of MSL-N medium[32]. Cells were then diluted in 3 ml of MSL-N medium to final $OD_{600} = 1.5$ and incubated at 30 °C with 200 rpm agitation for 4–6 h. Finally, cells were mounted onto MSL-N agarose pads, covered with a coverslip and the chamber was sealed using VALAP (vaseline, lanolin and paraffin; 1 : 1 : 1).

For flow cytometry and quantifications of mating and sporulation efficiencies, and mating and sporulation defects, ~3×10^7 of MSL-N washed cells were re-suspended in 20 μl of MSL-N medium and spotted onto MSL-N 2% agar plates and incubated at 25 °C for the indicated number of hours. The number of unmated cells, unsporulated zygotes and sporulated zygotes was determined using transmitted light microscopy, and mating and sporulation efficiencies were calculated using the following formulas:

$$\text{Mating efficiency (\%)} = \frac{2n_{\text{zygotes}}}{n_{\text{unmated cells}} + 2n_{\text{zygotes}}} \times 100$$

$$\text{Sporulation efficiency (\%)} = \frac{n_{\text{sporulated zygotes}}}{n_{\text{zygotes}}} \times 100$$

The reported results are from three replicates, error bars denote standard deviation and *P* values were calculated using two-tailed Welch's *t*-test assuming normal distribution.

Diploid strains. The diploid strains were constructed from haploid P and M cells in which mating type switching was abolished due to a H1Δ17 mutation linked to natMX and kanMX selection cassettes, respectively. Cells of the two mating types were mated for 24 h on solid ME medium at 25 °C and then plated on YE solid medium containing both nourseothricin and kanamycin. Colonies resistant to both antibiotics were subjected to Hoechst staining followed by flow cytometry to identify diploid strains. Presence of both *mat1-P* and *mat1-M* loci was confirmed by diagnostic PCR.

Flow cytometry. We collected the mated cell suspensions from MSL-N plates and resuspended samples in 1 ml of MSL-N medium containing 10 μl of glusulase (NEE154001EA, Perkin-Elmer). After overnight incubation at 30 °C we visually verified that all cells except spores had lysed upon glusulase treatment. The samples were washed three times with MSL-N and resuspended in 3 ml of water containing 2–5 μg/ml of Hoechst 33342 DNA dye (B2261, Sigma-Aldrich) and incubated for 15–30 min at room temperature. Samples were immediately analysed on Becton Dickinson Fortessa with proprietor software platform using the 355-nm laser line with 450/50 filter. The experiment was run in duplicate.

Strains. All strains are reported in Supplementary Table 1. Standard fission yeast genetic and molecular biology tools were used in the study[33]. Sequences of mutants obtained in this study are provided as annotated, GenBank-formatted sequences available at https://figshare.com/s/6d81d779a8550b2caec6 (supplementary sequences 1–27).

The mating loci of strains used in the study are illustrated in Extended Data Fig. 6b. Mi and Pi deletions remove the ORF of the genes. sfGFP tagging of Pi and Mi was done at the N- and C- terminus, respectively.

In homothallic strains, the *mat1* locus determines the cell mating type and switches between sequences encoded by the silent *mat2* and *mat3* loci (reviewed in ref. [34]). Importantly, a H1Δ17 mutation at the *mat1* locus H1-homolgy box prevents mating type switching[35]. As depicted in the Extended Data Fig. 6b, in the synthetic *mat1* locus sequences, the wild-type H1-homology box was replaced with the one amplified from the PB9 strain carrying the H1Δ17 mutation (a gift from B. Arcangioli, Institut Pasteur). The recombinant DNA also contains a selection cassette and homology arms that targets the constructs to the *mat1* locus of homothallic cells. The described sequences were cloned and amplified inside the pBlueScript plasmid. Finally the indicated restriction enzymes (see supplementary sequences 1–4) were used to extract the fragment of interest and transform it into *h90* wild-type cells. Correct integration was verified by PCR and/or inability to perform mating type switching.

mat2 and *mat3* loci that reside in heterochromatic DNA and thus cannot be targeted by homologous recombination in wild-type cells were manipulated as described in ref. [36]. An overview of the experimental procedure is presented in Extended Data Fig. 7a. The *mat2* locus manipulations were performed in the TP1 strain (a gift from G. Thon, University of Copenhagen), which lacks the *clr3* gene, to allow recombination at the *mat2* locus. Furthermore, the TP1 strain carries *ade6+* and *ura4+* cassettes in the vicinity of the *mat2* locus. The TP1 strain was transformed with DNA fragments carrying the designed *mat2* locus and flanking wild-type sequences. 5-FOA resistant, adenine auxotrophs were selected and tested for correct integration by PCR. The *mat2* mutant locus was then introduced in otherwise wild-type cells via genetic crosses. *mat3* manipulations were similarly performed in the PG3089 strain (a gift from G. Thon), which lacks the *clr4* gene and carries the *ura4+* cassette in the proximity of the *mat3* locus. The sequences of the mutant *mat2* and *mat3* loci are provided (see supplementary sequences 5–8).

For Pi overexpression we used the constitutive *tdh1* promoter comprising 1000 bp upstream of the *tdh1* start codon to drive expression of Pi N-terminally tagged with sfGFP. This construct was cloned into plasmids that contain the budding yeast ADH1 transcriptional terminator sequence, BleMX selection cassette and sequences targeting integration at either *ade6* or *his5* locus (supplementary sequences 9, 10). PmeI digested *ade6* targeting plasmid and StuI digested *his5* targeting plasmid were transformed in *ade6* and *his5* mutant cells, respectively. Prototrophic, zeocin-resistant clones were selected and checked for correct integration by PCR and genetic crosses.

For Mi overexpression we used either the *tdh1* or the *nmt1* promoter to drive expression of Mi that was C-terminally tagged with either Halo tag or mCherry. These constructs were cloned into plasmids that target either *ura4* or *lys3* locus and contain kanMX or natMX selection cassettes, respectively (supplementary sequences 11, 12). AfeI digested *ura4* targeting plasmid and BstZ17I digested *lys3* targeting plasmid were transformed in corresponding auxotrophic strains and prototrophic clones with adequate antibiotic resistance were selected and checked for correct integration by PCR and genetic crosses.

*mei3*Δ mutant background was used to prevent induction of meiosis by co-expression of Mi and Pi[37,38] as was the case for control strains expressing individual proteins.

For Mi expression from the P-cell-specific p^{map3} promoter we used 2063 bp upstream of the *map3* start codon followed by the Mi ORF sequence, the budding yeast ADH1 transcriptional terminator sequence and BleMX selection. This construct was cloned inside the pJK148 plasmid[39] that targets the construct to the *leu1* locus after NdeI digestion (supplementary sequence 13). The linearized plasmid was transformed into *h−* cells lacking Mi, and a G418 resistant transformant that efficiently sporulated when crossed with wild-type *h+* cells was selected. The same transformation clone was used to introduce the *pak2*Δ*::hphMX* mutation as described below.

*pak2*Δ*::ura4+* mutant strain was derived from ref. [40]. For the *pak2*Δ*::hphMX* we cloned the 430 bp of 3′ UTR of *pak2* followed by 399 bp of the 5′ UTR of *pak2* into the pFA6a-hphMX vector, which created an AfeI site between the two UTRs. The obtained plasmid (supplementary sequence 14) was linearized with AfeI, transformed into cells and hygromycin-B resistant colonies were selected and tested for correct integration by PCR.

*mei2*Δ*::kanMX* and *mei3*Δ*::kanMX* mutant strains were derived from the *S. pombe* deletion library strains S7B02 and S20A08 respectively (Bioneer) and were verified by PCR.

*sme2*Δ*::ura4+* and *mei4*Δ*::ura4+* mutations were derived from Japan National BioResource Project strains FY7237 and FY7361, respectively.

*fus1*Δ*::natMX* was generated using methodology detailed in ref. [41]. In brief, primers with 80-bp homology to sequences immediately flanking the fus1 ORF (see Supplementary Table 3) were used to PCR amplify the natMX selection cassette from the pFA6a-natMX plasmid and transformed into cells. Nourseothricin resistant clones were selected and genotyped by PCR.

For construction of *mei2* point mutant we first cloned the 3′ region (sequence 313–672 bp downstream of *mei2* stop codon) followed by the *mei2* gene (sequences from 487 bp upstream of the start codon to 312 bp downstream of the stop codon), which creates an AfeI restriction enzyme between the two fragments. This construct was inserted in the pFA6a-hphMX plasmid and we introduced the F644A mutation by site-directed mutagenesis (supplementary sequences 15, 16). Linearization of the plasmid with AfeI targets the construct to replace the genomic *mei2* gene upon transformation. The hygromycin-B resistant clones were selected and the correct integration replacing the wild-type *mei2* was confirmed by PCR.

The strains with a fluorescently labelled genomic locus[42] were derived from FY13708 strain obtained from the Japan National BioResource Project. In these cells the *lys1+* locus is genetically linked to the *LacO* sequence array that binds the GFP–LacI expressed from the *dis1* promoter at the *his7+* locus.

For microtubule visualization we used cells with SV40 promoter driven expression of GFP–α-tubulin[43] derived from a FC1234 strain, a gift from F. Chang's group.

For *mei3* fluorophore tagging, yeast codon optimized sfGFP (a gift from M. Knop, Heidelberg University) was introduced into a pFA6a-natMX vector to obtain the pFA6a-sfGFP-natMX vector (supplementary sequence 17). The plasmid was then used as a template for PCR with primers that amplify sfGFP followed by natMX cassette and carry 80-bp homology to sequences immediately flanking *mei3* STOP codon (see Supplementary Table 2). The PCR product was transformed into wild-type cells, and nourseothricin resistant clones were selected and genotyped by PCR. All *mei3*–sfGFP expressing strains used in the study were derived from the same transformation clone by genetic crosses.

For C-terminal mCherry tagging of nuclear marker *uch2* and spindle pole body marker *pcp1*, we used the pFA6a-mCherry-natMX and pFA6a-mCherry-kanMX plasmids respectively. In both cases, we cloned the 3′ UTR region of the gene followed by C-terminal fragment of the ORF keeping it in frame with the fluorophore (supplementary sequences 18, 19). The obtained plasmids to tag *uch2* and *pcp1* were then treated with AfeI and AfeI + SnaBI restriction enzymes respectively, and the digested DNA was transformed into cells. Clones resistant to either nourseothricin or G418 were selected and genotyped by PCR.

myo52–GFP and myo52–tdTomato were introduced from strains FC857[44] and ySM740[45], respectively.

The construct to express the green and red fluorophores from the P-cell-specific *map3* promoter were previously described in ref. [46]. For the expression of GFP from the M-cell-specific *mam2* promoter we cloned the 438 bp of promoter upstream of the start site into a pRIP-GFP plasmid that contains the *ura4*+ selection cassette (supplementary sequence 20). Transformation of the *ura4*-mutant cells with the PmeI digested final plasmid targeted the construct to the native *mam2* promoter and conferred growth in absence of uracil.

Sequences encoding small interacting SynZip peptides[47] were a gift from S. Pelet, University of Lausanne. The SynZip3 and SynZip4 peptides were cloned at the C terminus of mCherry and sfGFP, respectively. Expression of mCherry–SynZip3 was driven from the p^{act1} promoter consisting of 822 bp upstream of the *act1* start codon. The sfGFP–SynZip4 was driven from the p^{tdh1*} promoter comprising 896 bp upstream of the *tdh1* start codon. For the NLS–sfGFP–SynZip4 we introduced the SV40 nuclear localization signal at the sfGFP N terminus and used the p^{tdh1*} promoter comprising 896 bp upstream of the *tdh1* start codon. All constructs were integrated in the plasmid that carries the *ura4*+ selection cassette and targets the *ura4* locus. All constructs are detailed in supplementary sequences 21–23. The AfeI linearized final plasmids were transformed into *ura4* mutants and uracil prototrophs were selected for further experiments.

To replace the *mei3* coding sequence with mCherry we designed a plasmid that upon restriction digestion with PacI and SphI produced a linear fragment in which 405 bp upstream of *mei3* start codon were fused to mCherry sequence and pFA6a-hphMX backbone sequence followed by 104 bp downstream the *mei3* stop codon. This fragment (supplementary sequence 24) was purified, transformed into wild-type cells, and hygromycin-B resistant clones were selected. Correct integration replacing the wild-type mei3 was confirmed by PCR.

To simultaneously label mating M cells with green fluorescence and P cells with red fluorescence, we cloned into a plasmid to target the *ade6* locus (supplementary sequence 27): 1,751 bp of the *mam1* promoter, with a point mutation in the AfeI site, in front of sfGFP coding sequence followed by budding yeast ADH1 transcriptional terminator, the blasticidin-S resistance gene, and 2,063 bp of *map3* promoter driving expression of mCherry fused with another copy of the budding yeast ADH1 transcriptional terminator. PmeI linearized plasmid was transformed into cells, blasticidin-S resistant clones were selected and correct integration confirmed by PCR.

For DNA extraction to check efficiency of nuclease treatment, cell lysates were mixed with sodium-dodecyl-sulfate (1%, final concentration) and either sodium acetate (300 mM, final concentration) or lithium acetate (250 mM, final concentration) and incubated at 70 °C for 10 min. After adding two volumes of ethanol and vortexing, the DNA was pelleted by centrifugation at 4 °C for 15 min at 13,000*g*. Pellets were washed with 1 ml of 70% ethanol, air-dried, re-suspended in TE buffer and incubated at 50 °C for one hour. Samples were analysed on agarose gels containing SYBR Safe (S33102, Thermo Fisher).

Protease inhibitors mix used in biochemical purifications included 100 mM PMSF, 15.6 μg/ml benzamidine-HCl, 0.5 μg/ml antipain, 0.5 μg/ml aprotinin, 0.5 μg/ml chymostatin, 0.5 μg/ml leupetin, 0.5 μg/ml pepstatin A, 0.5 μg/ml phenanthroline and a commercial inhibitor cocktail (11836153001, Roche)

Co-immunoprecipitation. We grew 100 ml of *mei3*Δ mutant cell cultures overexpressing Mi–Halo or sfGFP–Pi or both to exponential phase and collected cells by centrifugation for one minute at 1,000 rcf. Subsequent steps were performed on ice unless otherwise indicated. Cells were transferred to 2-ml microcentrifuge tubes, washed twice with ice-cold PBS buffer (NaCl 137 mM, KCl 2.7 mM, Na_2HPO_4 10 mM, KH_2PO_4 1.8 mM, pH 7.5, protease inhibitors mix) and re-suspended in 500 μl of the PBS buffer and ~1 ml of acid-washed glass beads were added. Cells were lysed using the FastPrep-24 bead beater (MP Biomedicals) set to 4.5 m/s with 10 cycles of 20-s beating and 40-s cooling on ice. The beads were removed and samples transferred to a new tube in which the cell debris was pelleted by centrifugation for 15 min at 13,000*g*. We collected the supernatant and determined the protein concentration using the Bradford assay. We adjusted the samples to 1.5 mg of total protein in 800 μl of PBS buffer, added 1 μl of TMR-Halo Ligand (G8251, Promega) and kept samples protected from light hereafter. Each sample was split in half and one aliquot was treated with 2 kU of benzonase (E1014, Sigma-Aldrich) for one hour at room temperature on a rotator stand. Twenty microlitres of the sample was used for DNA extraction (see below) and 40 μl for western blotting analysis (Fig. 3d labelled as input). The samples were then incubated with 10 μl of GFP-Trap beads (gtma-20, Chromotek) prewashed in PBS buffer. After one hour incubation on a rotating stand at 4 °C beads were washed two times in PBS buffer and two times in CXS buffer (50 mM HEPES, 20 mM KCl; 2 mM EDTA; 1 mM $MgCl_2$, pH 7.5; protease inhibitors) with each wash performed in a new microcentrifuge tube. Beads were then re-suspended in Laemmli buffer, heated to 70 °C for 10 min and then subjected to SDS–PAGE (XP10205BOX, Thermo Fisher). Mi–Halo signal was recorded using the Cy3 channel on the Amersham Typhoon 5 system with proprietor software (General Electric) and the gel was then blotted onto nitrocellulose membrane using the Towbin transfer buffer (25 mM Tris-base, 192 mM glycin, 20% methanol, pH 8.3). After blocking in 5% milk, membrane was first probed with anti-GFP antibody (Cat.No. 11814460001, Roche; 1:3,000 dilution) followed by secondary HRP-coupled anti-mouse antibody (W402B, Promega; 1:3,000 dilution), developed and visualized on the Fusion FX (Vilber). Subsequently, the same membrane was probed with the TAT-1 antibody (a gift from K. Gull, University of Oxford) targeting tubulin as a loading control. DNA was extracted as described using lithium acetate and analysed on agarose gel (Extended Data Fig. 8a). The experiment was reproduced in three independent replicates.

Co-purification of recombinant Pi and Mi. Purification of Pi N-terminally tagged with either MBP, GST or sfGFP all failed owing to the instability of the protein in both *E. coli* cells and post-purification (data not shown). However, co-expression of Mi greatly stabilized Pi and thus we proceeded to purify the recombinant Pi–Mi complexes. The expression plasmids used (supplementary sequences 25, 26) are derived from the pMAL backbone and use the inducible *lacIq* promoter to drive expression of bicistronic mRNA encoding MBP–Pi followed by two stop codons, an internal ribosomal entry site (IRES) and either Mi or Mi–6×His. Western blotting showed that there was no considerable translational read-through between the ORFs of Pi and Mi. Coding sequences of both Pi and Mi were codon optimized for expression in *E. coli*.

An *E. coli* BL-21 strain carrying the expression vector was inoculated in 25 ml of LB medium with 100 μg/ml ampicillin and incubated overnight at 37 °C with shaking at 200 rpm. The next morning cultures were diluted in 25 ml of medium to $OD_{600} = 0.2$ and allowed to grow to $OD_{600} = 0.8$ before diluting them again in 150 ml of medium to $OD_{600} = 0.1$ and incubated at 37 °C with shaking until OD_{600} reached 0.6. At this point isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to final concentration 150 μM and cells incubated for 6 h at 18 °C with 200 rpm shaking. Cells were then collected by centrifugation at 3,000 rcf for 15 min, washed in 25 ml of buffer L (30% glycerol v/v, 50 mM Tris-HCL pH 7.5, 100 mM KCl, 1 mM imidazole) and cell pellets frozen in liquid nitrogen.

All subsequent steps were performed in presence of the protease inhibitor mix described above. Frozen cell pellets were thawed on ice and re-suspended in 8 ml of buffer LPI (50 mM Tris-HCL pH 7.5, 100 mM KCl, 1 mM imidazole, protease inhibitors mix) and sonicated on ice with Sonoplus HD2070 (Bandelin) at 45% power output for 5 cycles of 30-s pulses followed by 30-s pauses. Lysates were shifted to microcentrifuge tubes and cell debris removed by centrifugation at 15,000 rcf for 15 min at 4 °C. Lysates were then pooled together in a 15-ml falcon tube and Bradford assay was used to measure protein content. Protein concentrations of all samples were equalized with buffer LPI, 100-μl aliquots saved for DNA extraction and the remaining sample treated with 2 kU of benzonase for two hours at room temperature with rotation. Another set of 100-μl aliquots were collected to extract DNA (Extended Data Fig. 8b) and 100-μl aliquots were set aside for probing by western blotting (Fig. 2e, input). Each sample was then mixed with 200 μl of HisPur cobalt resin (89964, Thermo Fisher) prewashed in buffer LPI and samples were incubated with rotation at 4 °C for 4 h. Beads were pelleted by centrifugation at 600 rcf for 2 min at 4 °C and washed twice with 10 ml of buffer W1 (30% glycerol v/v, 50 mM Tris-HCL pH 7.5, 100 mM KCl, 10 mM imidazole) and twice more with 10 ml of buffer W2 (30% glycerol v/v, 400 mM NaCl, 50 mM Tris-HCL pH 7.5, 100 mM KCl, 10 mM imidazole). For elution the beads were incubated with buffer E (30% glycerol v/v, 50 mM Tris-HCL pH 7.5, 100 mM KCl, 250 mM imidazole). Eluates were then re-suspended in Laemmli buffer and used for SDS–PAGE with proteins visualized through Coomassie staining and western blotting. For primary antibodies we used 1:4,000 dilution of anti-MBP antibody (2396, Cell Signaling) and 1:2,000 anti-His antibody (34660, Qiagen). Infrared fluorophore coupled secondary antibodies at 1:500 dilution (R-05061 and R05054, Advansta) were used and blots visualized with the Fusion

FX (Vilber). DNA was extracted as described above using sodium acetate and analysed on agarose gel (Extended Data Fig. 8b). The experiment was performed in biological triplicate.

Chromatin immunoprecipitation. Mating between fission yeast cells is asynchronous and thus we performed the chromatin immunoprecipitation on diploid cells. Because Pi and Mi induce *mei3* only in early meiosis, we prevented meiotic progression by replacing the *mei3* ORF with that of mCherry. These adjustments resulted in the majority of cells responding to nitrogen starvation by strongly inducing mCherry expression from the *mei3* promoter in a Pi- and Mi-dependent manner (data not shown). We used cells expressing untagged Mi and Pi to determine the background DNA binding. The ChIP testing Mi binding to the *mei3* promoter was performed on diploid cells carrying Mi–sfGFP at the *mat1-M* locus while the *mat1-P* locus either encoded wild-type Pi or lacked its ORF. Likewise, to test Pi binding to the *mei3* promoter we used diploid cells expressing sfGFP–Pi from the *mat1-P* locus while the *mat1-M* locus either encoded wild-type Mi or lacked its ORF.

The experiment was performed in biological triplicate. Cells were grown at 30 °C overnight to exponential phase in 50 ml of EMM2 medium[33] and then diluted in 100 ml of EMM2 medium to a final $OD_{600} = 0.03$. After 12 h incubation, cells reached exponential growth ($OD_{600} \approx 0.6$) and were collected by centrifugation at 1,000 rcf for 3 min. Cell pellets were then washed 3 times with 20 ml of EMM-N medium (EMM2 lacking ammonium chloride) and re-suspended in 50 ml of EMM-N to final $OD_{600} = 1$. The suspensions were incubated for 10 h at 25 °C and shifted to 50-ml centrifugation tubes. To cross-link proteins and DNA, we added 1.3 ml of 37% formaldehyde solution (F8775, Sigma-Aldrich) and incubated at room temperature for 15 min on a tube roller. Cross-linking was quenched with 2.6 ml of 2.5 M glycine followed by 5 min incubation at room temperature on a tube roller. Cells were then collected by centrifugation at 1,000 rcf for 3 min at 4 °C, washed twice with 15 ml of ice-cold PBS buffer, transferred to 2-ml screw-cap tubes and cell pellets frozen in liquid nitrogen for 30 min. Next, samples were thawed on ice and resuspended in 500 µl of freshly prepared lysis buffer (50 mM Hepes/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% TritonX-100 0.1% Na-deoxycholate, 1 mM PMSF and inhibitor cocktail (11836153001, Roche)). After adding 0.5-mm zirconium beads nearly to the meniscus, cells were lysed using the FastPrep-24 5G bead beater (MP Biomedicals) at 6.5 m/s with 3 cycles of 60-s beating and 5 min cooling on ice. Crude lysates were transferred to 15-ml polystyrene falcon tubes, and the sample volume was adjusted to ~1.5 ml before sonification using Bioruptor Plus system equipped with a 4 °C water-bath (Diagenode). We performed 3 rounds of sonication: Each sonication round consisted of 10 cycles of 30-s high-intensity sonication, 30-s pause and was followed by one 10-min pause on ice. Samples were shifted to microcentrifuge tubes and cell debris pelleted by centrifugation at 13,000 rcf for 5 min at 4 °C. Supernatants were transferred to new microcentrifuge tubes and cleared by 13,000 rcf for 15 min at 4 °C. The DNA concentration was measured in each lysate by QuBit fluorometric quantitation (Invitrogen) and all lysates adjusted to same DNA concentration. The samples were then incubated with rotation for 8 h at 4 °C with 30 µl of magnetic GFP-Trap beads (gtma-20, Chromotek) prewashed three times in PBS buffer with 0.02% Tween. Next, beads were pelleted by centrifugation at 100 rcf for 1 min at 4 °C and immobilized using the magnetic rack before discarding the supernatant. Beads were then subjected to three washes with 1 ml of freshly prepared ice-cold lysis buffer, one wash with 1 ml of freshly prepared ice-cold wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA) and one wash with 1 ml of ice-cold TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Each wash was performed at room temperature with tube rotation that lasted 5 min. Beads were eluted by incubation with 125 µl of TES buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% SDS) at 65 °C for 10 min with agitation. Supernatant was collected and the beads eluted once more with 125 µl of TES buffer. The two elutes were pooled and incubated overnight at 65 °C with agitation to reverse the formaldehyde-induced crosslinking. Samples were cooled down and then incubated with 4 µl of 10 mg/ml RNase A (10109169001, Roche) for 1 h at 37 °C and subsequently with 3 µl of 20 µg/ml proteinase K (3115879001, Roche) for 1 h more at 37 °C. We added 15 µl of 5 M NaCl, 300 µl of isopropanol and 30 µl of DNA-binding AMPure XP beads (Beckman Coulter) to each sample and incubated samples at room temperature for 15 min with rotation. Beads were pelleted by centrifugation at 100 rcf for 1 min at 4 °C, immobilized using the magnetic rack and washed twice with 500 µl of 80% ethanol. The beads were briefly dried and DNA eluted with 20 µl of Tris-HCl pH 8.0. Obtained DNA was diluted twofold and used as template in real-time PCR with primers at final concentration 0.3 µM and SsoAdvanced Universal SYBR Green Supermix (172-5274 Bio-Rad) performed on 7500 Fast Real-Time PCR System (Applied Biosystems). Primers used in the qPCR are listed in the Supplementary Table 3). The qPCR was done in duplicate.

Data normalization was performed based on actin promoter binding and the binding observed in cells that do not express the affinity tag is considered as unspecific and is set to 1. Thus, fold enrichment for the locus of interest (loi) was calculated as

$$2^{\Delta C_t^{\text{sample}} - \Delta C_t^{\text{background}}}$$

in which $\Delta C_t = C_t^{Pact1} - \Delta C_t^{\text{loi}}$ and C_t is the threshold cycle. Reported values for each of biological triplicates are averages of two technical replicates. Position of the middle nucleotide of the qPCR amplicon in reference to the start codon was used in the graph. Because the amplitude of binding of the Pi–Mi complex to the *mei3* promoter varied between biological replicates, we present the data separately.

Microscopy and image analysis. Imaging of mating cells was performed as described[31]. We note that the frequency of re-fertilization events in *mei3*Δ, *Pi*Δ and *Mi*Δ strains appeared to increase further with time than the numbers reported but this increase could not be quantified owing to deterioration of the imaging chambers.

To acquire images in Fig. 3c and Extended Data Fig. 1c we used a spinning-disk confocal system that uses an inverted microscope (DMI6000B; Leica) equipped with an HCX Plan Apochromat 100×/1.46 NA oil objective and an UltraVIEW system (PerkinElmer; including a real-time confocal scanning head (CSU22; Yokagawa Electric Corporation), solid-state laser lines, and an electron-multiplying charge coupled device camera (C9100; Hamamatsu Photonics)). Stacks of *z*-series confocal sections were acquired at 0.3-µm intervals using Volocity software (Perkin Elmer).

All other micrographs were obtained by wide-field microscopy performed on a DeltaVision platform (Applied Precision) composed of a customized inverted microscope (IX-71; Olympus), a UPlan Apochromat 100×/1.4 NA oil objective, a camera (CoolSNAP HQ2; Photometrics), and a colour combined unit illuminator (Insight SSI 7; Social Science Insights). Images were acquired using softWoRx v.4.1.2 software (Applied Precision). Principally we imaged a single *z*-section with the exception of data presented in Fig. 1b and Extended Data Fig. 1b in which 6 sections with 0.5-µm spacing were acquired and the presented images are a maximum projection of images deconvolved using softWoRx v.4.1.2 inbuilt module. All conclusions based on imaging data were derived from at least three independent observations.

Image processing and fluorescence intensity based quantifications were performed in ImageJ (NIH). Supplementary Videos were converted from TIFF to MOV format using the inbuilt MPEG-4 compression. For overnight time-lapses that exhibited drifting, presented images were aligned using the MultiStackReg1.45 Plugin. All quantifications were performed on raw images.

The lifetime of the fusion focus was measured on time-lapse images as the interval between the fluorescently labelled marker proteins first focalizing[46] and cell fusion. Results are reported with the box-and-whiskers plot in which centre lines show the medians, box limits indicate the 25th and 75th percentiles, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots and the Kruskal–Wallis *P* value is reported.

To distinguish the expression of *mei3* from P and M genome we performed crosses between heterothallic strains in which only one partner carried the *mei3-sfGFP* allele, and the other partner had the unlabelled, wild-type gene. Quantification of Mei3–sfGFP fluorescence induction used cell fusion as time zero, evidenced as the first time point with exchange of cytosolic RFP between partners. The mean signal in individual partners before fusion, and thus before *mei3* induction, was considered as background signal. Mean fluorescence of the whole zygote was recorded over time and the average of the indicated number of cells is reported with shaded regions denoting mean ± standard deviation. The post-fusion time for the mean zygotic intensity to reach 50 arbitrary units above background was recorded for each zygote and the results are reported with the box-and-whiskers plot in which centre lines show the medians, box limits indicate the 25th and 75th percentiles, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots and the Kruskal–Wallis *P* value is reported.

For Mi–sfGFP nuclear entry dynamics we used the last time-point before exchange of Mi–sfGFP between partners as time zero. The nuclei of the zygote were clearly identifiable based on Mi–sfGFP signal alone, which enabled us to outline the nuclear region required in these quantifications. The signal in the P cell before cell fusion was considered as background. Mean fluorescence of each nuclear region was recorded over time and the average of the indicated number of cells is reported with shaded regions denoting mean ± standard deviation and the Kruskal–Wallis *P* value is reported.

For quantification of dynamics of NLS–sfGFP–SynZip4, sfGFP–SynZip4 and mCherry–SynZip3 we used the last time-point before exchange of red fluorophore between partners as time zero. The nuclear region was outlined on the basis of the accumulation of the fluorescent signals in the central cell region. The signal in the partner cell lacking the fluorophore before cell fusion was considered as background signal. Mean fluorescence of each nuclear region was recorded over

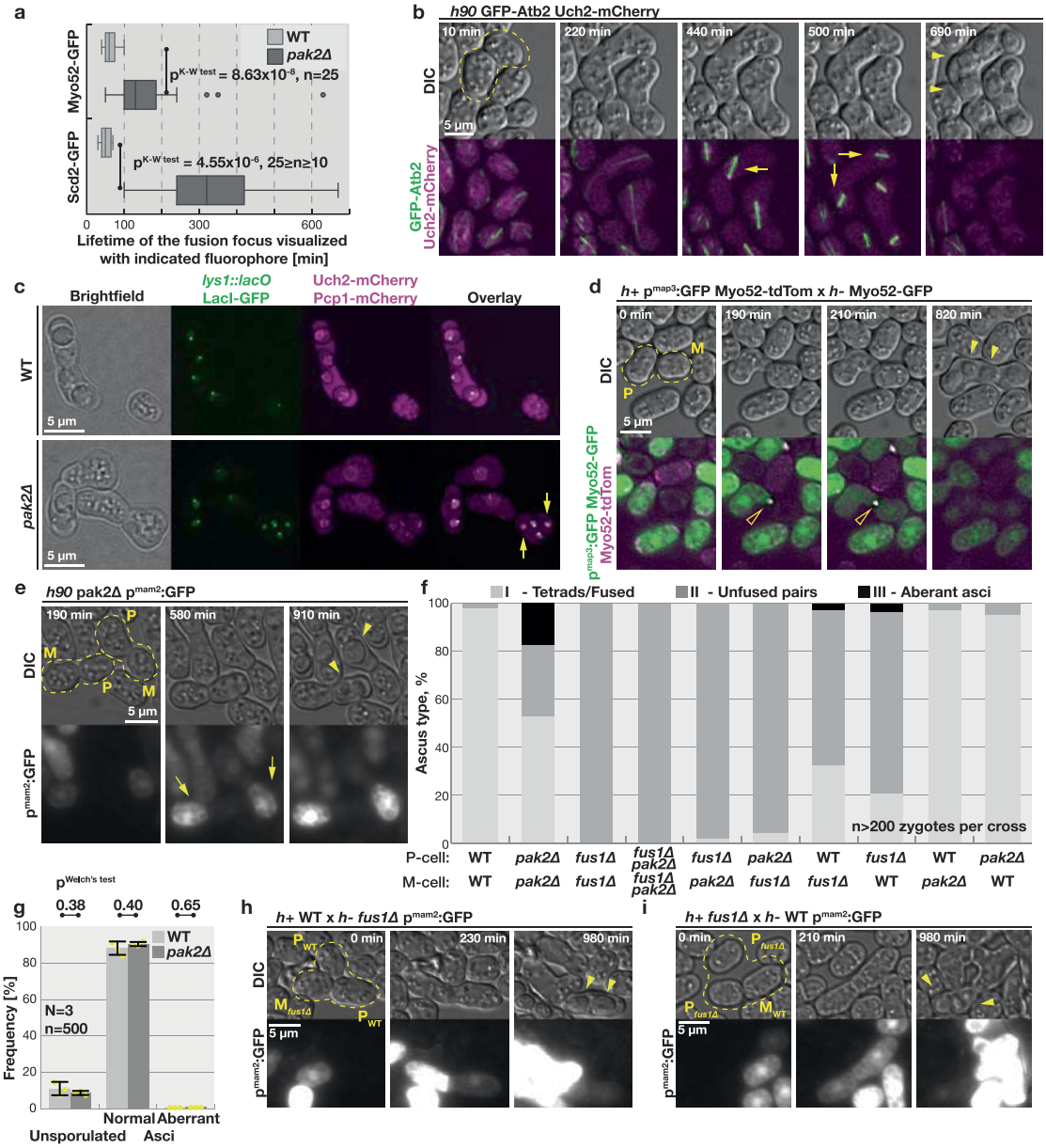
time and the average of the indicated number of cells is reported with shaded regions denoting mean ± standard deviation and the Kruskal–Wallis *P* value is reported. Note that the mCherry–SynZip3-expressing cell also displays background vacuolar fluorescence, probably owing to pre-fusion degradation of some mCherry–SynZip3.

Statistical analysis. Exact sample size (*n*) is reported for each quantification. Conclusions were drawn on experimental triplicates, except for Extended Data Fig. 1f, which was performed quantitatively only once (but qualitatively several times), while Fig. 1c and Extended Data Fig. 2h were done twice. No cells were excluded from the analyses, but zones of multi-layering were excluded from quantifications because individual cells could not be tracked reliably. Measurements were taken from distinct samples. To calculate *P* values for samples we assumed followed normal distribution, we used the two-tailed Welch's test calculated by Excel T.TEST() function. We also report Kruskal–Wallis *P* values ($p^{K-W\ test}$) that were obtained by Matlab ranksum() function set for two-sided comparisons of samples. All *P* values are given as exact values except for Extended Data Fig. 4b in which *P* values are greater than 0.05. No Bayesian analysis was performed. No hierarchical and complex designs are present. No explicit estimates of effect sizes have been performed, yet all data are shown in absolute or normalized values so effect sizes are evident to the reader.

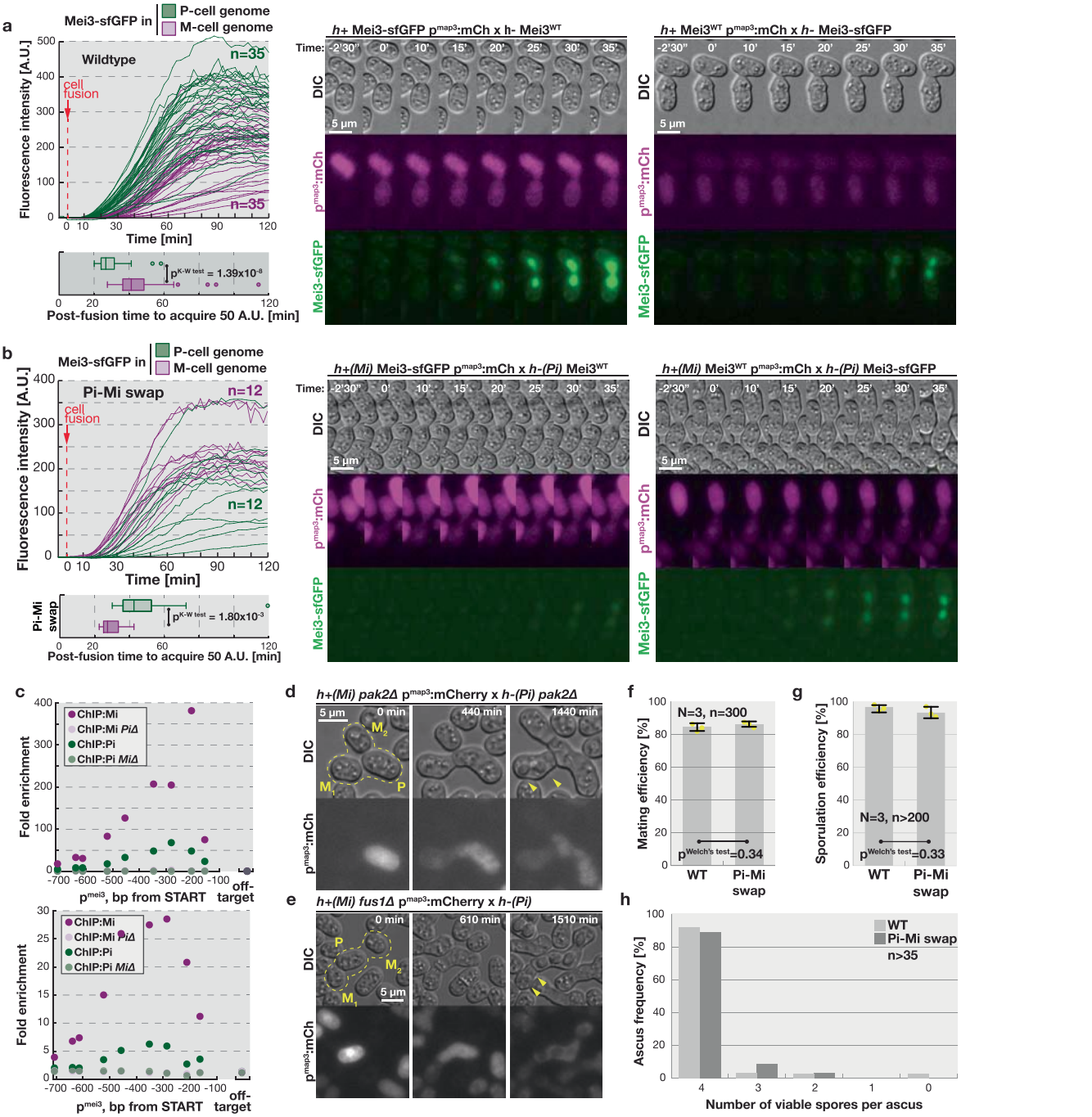
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

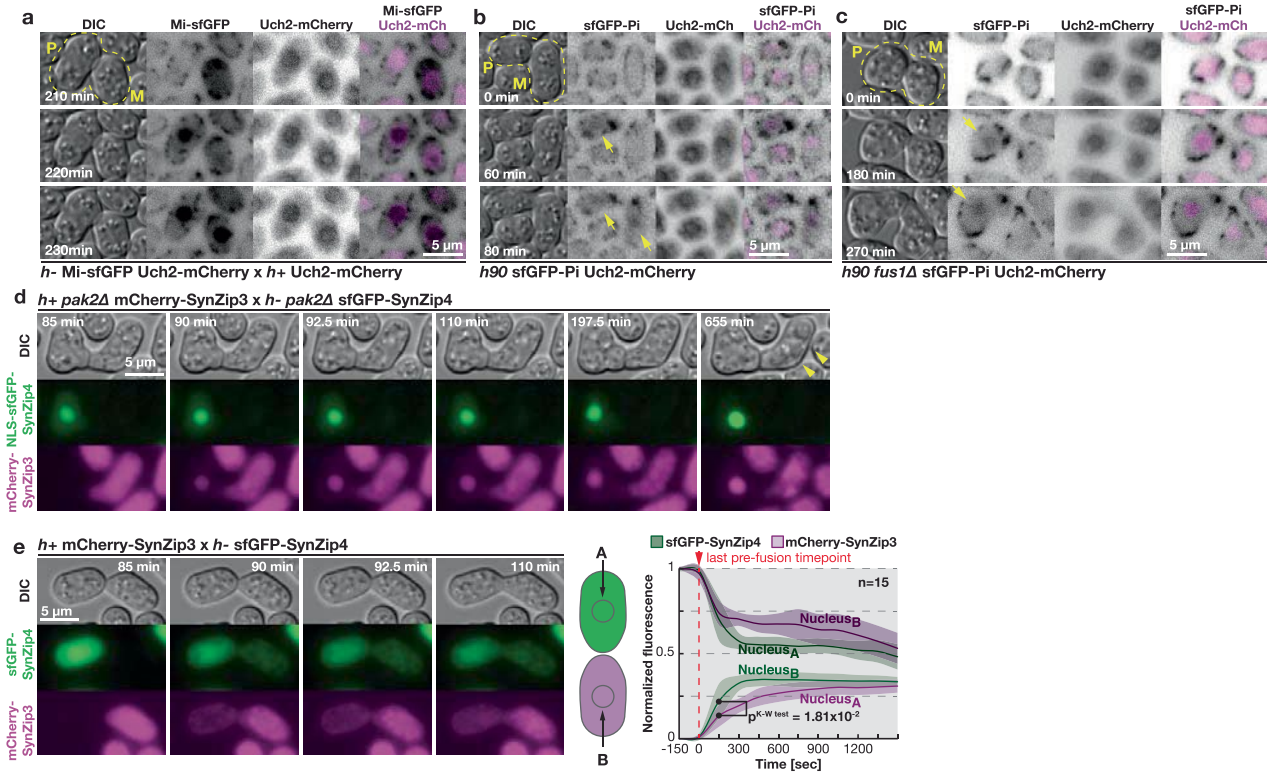
31. Vjestica, A., Merlini, L., Dudin, O., Bendezu, F. O. & Martin, S. G. Microscopy of fission yeast sexual lifecycle. *J. Vis. Exp.* **109**, (2016).
32. Egel, R., Willer, M., Kjaerulff, S., Davey, J. & Nielsen, O. Assessment of pheromone production and response in fission yeast by a halo test of induced sporulation. *Yeast* **10**, 1347–1354 (1994).
33. Hagan, I. M., Carr, A. M., Grallert, A. & Nurse, P. (eds). *Fission Yeast: A Laboratory Manual* (CSH Press, Cold Spring Harbor, 2016).
34. Klar, A. J. S. Lessons learned from studies of fission yeast mating-type switching and silencing. *Annu. Rev. Genet.* **41**, 213–236 (2007).
35. Arcangioli, B. & Klar, A. J. A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient mat1 switching in Schizosaccharomyces pombe. *EMBO J.* **10**, 3025–3032 (1991).
36. Jakočiūnas, T., Holm, L. R., Verhein-Hansen, J., Trusina, A. & Thon, G. Two portable recombination enhancers direct donor choice in fission yeast heterochromatin. *PLoS Genet.* **9**, e1003762 (2013).
37. Willer, M. et al. Two-step activation of meiosis by the mat1 locus in Schizosaccharomyces pombe. *Mol. Cell. Biol.* **15**, 4964–4970 (1995).
38. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. Four mating-type genes control sexual differentiation in the fission yeast. *EMBO J.* **7**, 1537–1547 (1988).
39. Keeney, J. B. & Boeke, J. D. Efficient targeted integration at leu1-32 and ura4-294 in Schizosaccharomyces pombe. *Genetics* **136**, 849–856 (1994).
40. Bimbó, A. et al. Systematic deletion analysis of fission yeast protein kinases. *Eukaryot. Cell* **4**, 799–813 (2005).
41. Bähler, J. et al. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**, 943–951 (1998).
42. Nabeshima, K. et al. Dynamics of centromeres during metaphase–anaphase transition in fission yeast: Dis1 is implicated in force balance in metaphase bipolar spindle. *Mol. Biol. Cell* **9**, 3211–3225 (1998).
43. Minc, N., Boudaoud, A. & Chang, F. Mechanical forces of fission yeast growth. *Curr. Biol.* **19**, 1096–1101 (2009).
44. Feierbach, B. & Chang, F. Roles of the fission yeast formin for3p in cell polarity, actin cable formation and symmetric cell division. *Curr. Biol.* **11**, 1656–1665 (2001).
45. Martin, S. G., Rincón, S. A., Basu, R., Pérez, P. & Chang, F. Regulation of the formin for3p by cdc42p and bud6p. *Mol. Biol. Cell* **18**, 4155–4167 (2007).
46. Dudin, O. et al. A formin-nucleated actin aster concentrates cell wall hydrolases for cell fusion in fission yeast. *J. Cell Biol.* **208**, 897–911 (2015).
47. Reinke, A. W., Grant, R. A. & Keating, A. E. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* **132**, 6025–6031 (2010).



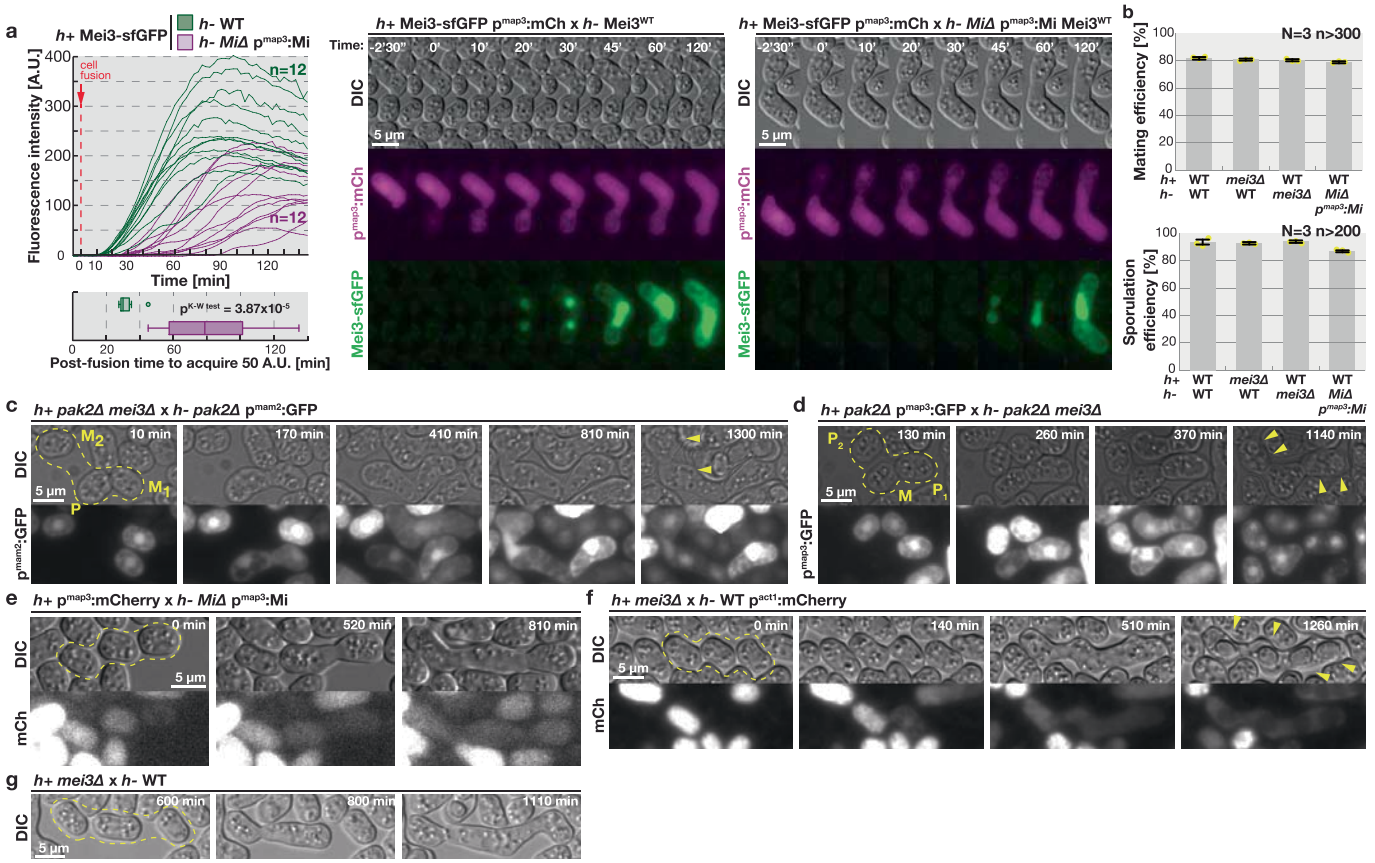
Extended Data Fig. 1 | Transient cell fusion results in ectopic meiosis in the P cell. a, Box-and-whiskers plot (see Methods) reports the lifetime of the fusion focus visualized by the indicated fluorophores in homothallic wild-type and *pak2*Δ mating cells. Kruskal–Wallis *P* value is reported. **b**, Wild-type control for data presented in Fig. 1b. Time-lapse of homothallic wild-type mating cells expressing GFP–α-tubulin in green and nuclear marker Uch2–mCherry in magenta. Note karyogamy in the outlined mating pair, meiotic spindles (arrows) and spores (arrowhead). **c**, Micrographs of cells in mating mixtures of homothallic wild-type and *pak2*Δ strains with *lys1* chromosomal loci labelled by LacO:GFP–NLS–LacI system in green and nuclei visualized by Uch2–mCherry in magenta. Arrows point to spores that lack the *lys1* locus. **d**, Wild-type control for data presented in Fig. 1d. Note the formation of the fusion focus (empty arrowhead), labelled by Myo52–GFP in the M cell and Myo52–tdTomato in the P cell, followed by exchange of cytosolic GFP expressed from the P-cell-specific p^{map3} promoter and spore formation (full arrowheads). **e**, Time-lapse showing mating of homothallic *pak2*Δ cells expressing GFP from the M-cell-specific p^{mam2} promoter. Note that the GFP exchange is followed by fusion pore closure and build-up of the GFP signal only in the M cells (arrows), whereas P cells form spores (arrowheads). **f**, P and M cells with indicated genotypes were induced to mate on solid medium. Two days later, we quantified frequencies of phenotypes indicated in the insets of Fig. 1a, $n > 200$. **g**, Quantification of sporulation phenotypes of wild-type and *pak2*Δ diploid cells upon nitrogen starvation, $n = 3$ experiments with $n = 500$ cells each. **h**, Time-lapse showing transient fusion between wild-type P cells and GFP-expressing *fus1*Δ M cells. Persistent GFP signal difference between partners indicated fusion pore sealing. Note spore formation (arrowhead) only in the wild-type P cell. **i**, Time-lapse showing mating of *h* + *fus1*Δ mutant P cell and GFP-expressing wild-type M cells. Note that transient fusion between indicated partners is followed by resealing of the fusion pore, as visualized by continued accumulation of GFP only in the M cell, and sporulation (arrowheads) in the P cell.



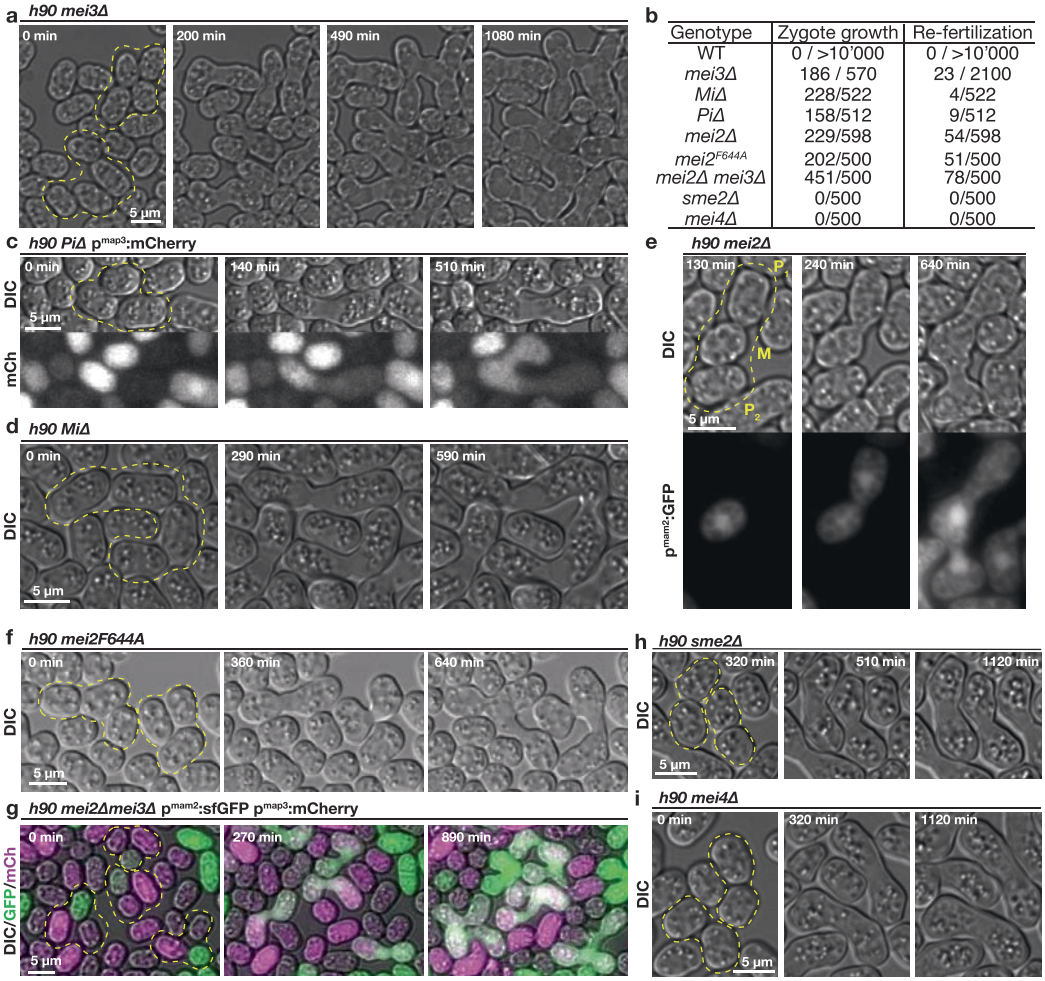
Extended Data Fig. 2 | Induction of the meiotic inducer Mei3 occurs first from the genome of the cell expressing the homeodomain protein Pi. a, **b**, Mei3–sfGFP signal increases more rapidly after fusion when expressed in the cell expressing Pi. The left panels show individual traces that quantify the fluorescent signal from zygotes of heterothallic wild-type (**a**) or Pi–Mi-swapped (**b**) cells in which only one partner has *mei3* tagged with sfGFP, as indicated (see Fig. 2c, g for average curves). Curves were aligned to fusion time defined by the entry into the M cell of cytosolic mCherry (expressed in the P cell). Box-and-whiskers plots in the bottom panels (see Methods) analyse data shown in the top panels, with Kruskal–Wallis *P* value displayed. Right panels show an example time-lapse of the cells used in the quantification. **c**, Biological replicates of chromatin immunoprecipitation of Pi and Mi reported in Fig. 2f. **d–e**, Time-lapse showing mating cells with swapped Pi and Mi coding sequences. In both cases, the P cell expresses cytosolic mCherry under control of p^{map3} promoter. **d**, Heterothallic *pak2*Δ strains, $n = 14$. **e**, Mating of otherwise wild-type M cells with *fus1*Δ P cells, $n = 5$. Transient fusion, visualized by florescence exchange, leads to haploid meiosis in the M_1 partner (arrowheads indicate spores). The P cell proceeds to mate (**e**) or attempts to mate (**d**) with the M_2 cell. **f**, **g**, Quantification of mating and sporulation efficiencies in heterothallic wild type and strains with swapped Pi and Mi coding sequences induced to mate on MSL-N agar plates for 28 h. **h**, Spores in asci produced by mating heterothallic wild type or strains with swapped Pi and Mi were micro-dissected and the number of spores that developed colonies were counted.



Extended Data Fig. 3 | Meiotic regulators Pi and Mi exhibit asymmetry in localization in early zygotes. a, Time-lapse showing mating of heterothallic cells expressing nuclear marker Uch2–mCherry (magenta) with the M cell co-expressing sfGFP-tagged endogenous Mi. Note the cytosolic Mi–sfGFP signal in M cells that rapidly accumulates in the P nucleus after partner fusion. The punctate signal at cell cortex in the green channel is background—probably mitochondrial—fluorescence. **b**, N-terminally sfGFP-tagged endogenous Pi exhibits a weak nuclear staining in P cells during mating in homothallic cells co-expressing nuclear marker Uch2–mCherry (magenta). Arrows point to nuclear sfGFP–Pi signal in the P cell (middle panel) and the zygote (bottom panel). **c**, Time-lapse showing homothallic, fusion-defective *fus1*Δ cells co-expressing N-terminally sfGFP-tagged endogenous Pi and nuclear marker Uch2–mCherry. Arrows point to Pi nuclear accumulation. **d**, Time-lapse showing transient fusion between *pak2*Δ cells expressing NLS–sfGFP–SynZip4 and mCherry–SynZip3. Note the accumulation of both fluorophores in the M nucleus and minimal transfer of the green fluorophore into the P cell that eventually sporulates (arrows) **e**, Left panel follows the mating of wild-type cells expressing cytosolic sfGFP–SynZip4 and mCherry–SynZip3. Right panel quantifies both fluorescent signals in the central region of the two cells corresponding to the two nuclei, as labelled on the scheme, and is presented as in Fig. 3e.

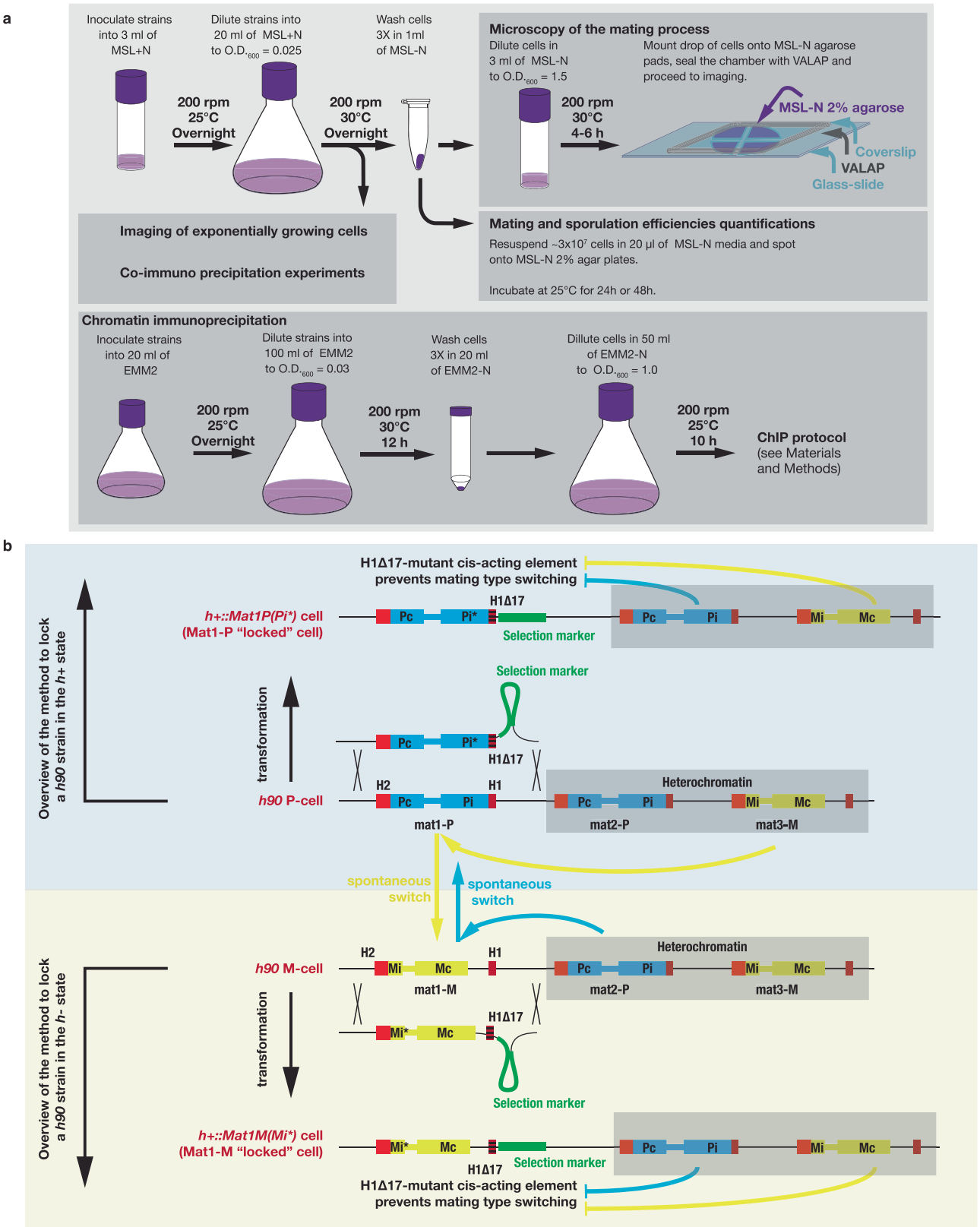


Extended Data Fig. 4 | Rapid induction of Mei3 is required to suppress mating in zygotes and prevent polyploid formation. a, Delayed Mei3–sfGFP signal when Mi expression is delayed. The left panel shows individual traces of the green fluorescent signal in zygotes produced by mating either wild-type or *MiΔ p^{map3}:Mi* M cells with P cells encoding Mei3–sfGFP. Curves were aligned to fusion time defined by entry into the M cell of cytosolic mCherry (expressed in the P cell). Box-and-whiskers plot in panel below (see Methods) analyses data shown in the top panel, with Kruskal–Wallis *P* value displayed. Right panels show representative time-lapse of cells used in the quantification. **b**, Heterothallic strains with indicated genotypes were mixed and induced to mate on MSL-N agar plates. Charts report mating and sporulation efficiencies quantified after two days incubation, $P > 0.05$ (Welch's test) for all comparisons to wild type, $n = 3$ experiments with $n > 200$ cells each. **c**, Mating of *pak2Δ* cells in which *mei3* has been deleted from the P partner and cytosolic GFP is expressed in M cells. Transient cytosolic exchange between P and M_1 cells never induced sporulation ($n = 10$). The partners shown continue mating with each other until the P cell eventually switches partner. **d**, Time-lapse shows *pak2Δ* P cells expressing cytosolic GFP mating with *pak2Δmei3Δ* double mutant M cells. Note that transient fusion is followed by spore formation ($n > 10$) in both indicated P cells. Arrowheads point to spores. **e**, Time-lapse showing that mating of M cells with Mi expressed from the *p^{map3}* promoter with wild-type P cells results in zygotes undergoing fusion with additional partners. Fusion events are visualized through exchange of cytosolic mCherry (expressed in P cells). **f**, **g**, Time-lapse showing that zygotes undergo repeated fusion, evident from exchange of cytosolic mCherry (in **f**), in crosses in which *mei3* was specifically deleted in the P cells.

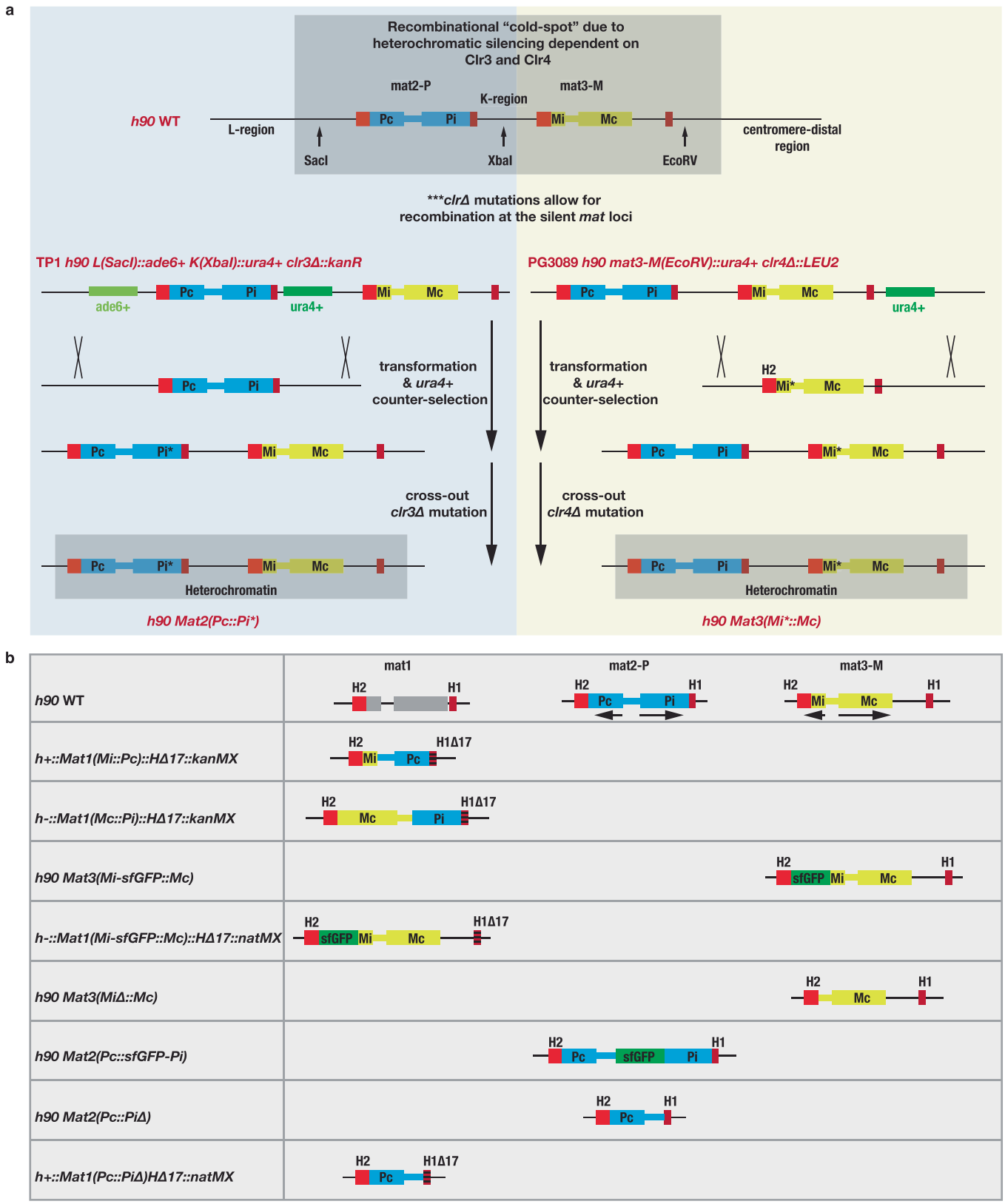


Genotype	Zygote growth	Re-fertilization
WT	0 / >10'000	0 / >10'000
mei3Δ	186 / 570	23 / 2100
MiΔ	228/522	4/522
PiΔ	158/512	9/512
mei2Δ	229/598	54/598
mei2^{F644A}	202/500	51/500
mei2Δ mei3Δ	451/500	78/500
sme2Δ	0/500	0/500
mei4Δ	0/500	0/500

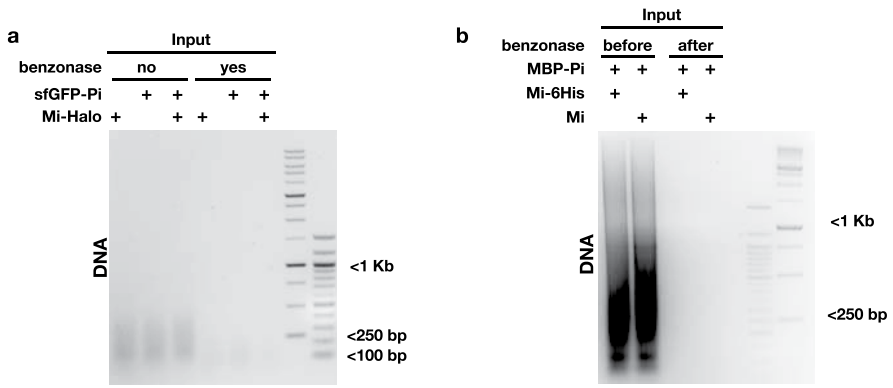
Extended Data Fig. 5 | Mei3–Mei2, but not sme2–Mei4, signalling suppresses mating in zygotes and prevents polyploid formation. **a**, Time-lapse showing re-fusion of *mei3*Δ homothallic cells. **b**, Incidence of re-fertilization and zygotic growth in mating mixtures of homothallic strains with indicated genotypes. **c**–**g**, Time-lapse showing re-fusion of *Pi*Δ, *Mi*Δ, *mei2*Δ, *mei2*^{F644A} and *mei2*Δ*mei3*Δ homothallic cells, as indicated. In **e**, fusion is evidenced through transfer of M-cell-expressed GFP. In **g**, fusion is evidenced through transfer of M-cell-expressed sfGFP and P-cell-expressed mCherry. **h**, **i**, Time-lapse showing diploid zygote formation in *sme2*Δ and *mei4*Δ homothallic cells, as indicated. Note that zygotes do not exhibit any growth or attempt mating with other cells.



Extended Data Fig. 6 | Schematic overview of growth conditions and *mat1* locus manipulations. a, Overview of growth conditions used in this study (adapted with permission from ref. [31]). See Methods for details. **b**, Schematic of *mat1* locus manipulations. The wild-type homothallic P cell and M cell *mat* loci are presented in the middle of the scheme. Red boxes denote H2 and H1 homology boxes identical between the three *mat* loci. Blue boxes represent *Pc* and *Pi* genes expressed exclusively in the P cell and encoded by the *mat2* locus, and yellow boxes denote *Mc* and *Mi* genes encoded by the *mat3* locus and expressed in M cells only. Gene expression from *mat2* and *mat3* loci is inhibited by heterochromatin state (grey box). The sequences at the *mat1* locus are derived from sequences at the other two *mat* loci during mating-type switching in homothallic strains (blue and yellow arrows), which relies on the H1 and H2 homology boxes. Transformation with a DNA fragment carrying the mutant H1Δ17 homology box (shaded red box) results in cells that are unable to switch mating type.



Extended Data Fig. 7 | Schematic overview of *mat2* and *mat3* loci manipulations and obtained *mat* loci mutants. a, Schematic of *mat2* and *mat3* loci manipulations. Wild-type homothallic strain is presented on the top. Red boxes denote H2 and H1 homology boxes identical between the three *mat* loci. Blue and yellow boxes represent genes encoded by the *mat2* and *mat3* loci, respectively. Recombination at the *mat2* and *mat3* loci is inhibited by heterochromatin state (grey box). Restriction enzyme sites denote positions at which prototrophic genes have been integrated in the TP1 and PG3089 strains depicted below, which are used to manipulate the *mat2* and *mat3* loci, respectively. Ablation of genes necessary for heterochromatin formation (*clr3* or *clr4*) enables targeting the *mat2* and *mat3* loci by homologous recombination. These deletions were subsequently crossed out. **b**, Schematic representation of *mat* loci mutants obtained in this study. Colour-coding is as in **a** and Extended Data Fig. 6b.



Extended Data Fig. 8 | Nuclease treatment removes majority of DNA from cell lysates. a, **b**, DNA from fission yeast (**a**) and bacterial (**b**) lysates treated with benzonase as indicated. These samples correspond to those shown in Fig. 2d, e, respectively.

VISIBILITY Female scientists deserve their own Wikipedia entries **go.nature.com/2nptwkq**

NATUREJOBS FACEBOOK For career info and job postings **www.facebook.com/naturejobs**

NATUREJOBS TWITTER For news, careers advice and jobs **twitter.com/naturejobs**

ADAPTED FROM GETTY

Teaming with bright ideas

Four researchers share their tips for building and maintaining international collaborations.

The number of international research collaborations continues to rise, and for good reason: it's easier than ever to connect with overseas colleagues, and doing so can be an effective way to share and advance knowledge. *Nature* spoke to four scientists who routinely participate in such teamings and who have studied how to create and look after them.

KATHRIN ZIPPEL
Be inclusive

Sociologist, Northeastern University, Boston, Massachusetts.

In our analysis of a 2006 US National Science Foundation survey of PhD holders in academic positions, we found that one-third of men reported taking part in an international collaboration, compared with only one-quarter of women (L. M. Frehill and K. Zippel *J. Wash. Acad. Sci.* **97**, 49–69; 2011) Women also publish less often with international colleagues than men do, according to a study that analysed research performance in 12 regions and 27 disciplines, over 20 years. Data from European studies show that women's international mobility starts to fall when they reach the postdoc level.

My recent book (K. Zippel *Women in Global Science: Advancing Academic Careers Through International Collaboration*; 2017) drew on interviews carried out between 2007 and 2015 with more than 100 women, many of whom were being held back by gender-specific 'glass fences' — organizational and structural barriers that control access to resources and opportunities.

It comes down to money and time. I found that funding is a major obstacle for women — whether for attending conferences to find collaborators or for travelling to establish collaborations. Women are often still the primary carers in their families, so travel comes with an extra financial burden in the form of care costs for children or other family members.

Women shouldn't hesitate to ask for what they need to scale a glass fence. For example, one woman I interviewed, who had four children, requested help to cover child care whenever she received an invitation to speak or work abroad. She went only if it was provided. Women can also turn to international funders, but those vary in terms of how they acknowledge caring responsibilities.

Women also tend to be in teaching roles, rather than in research-intensive ones, and are overburdened with service obligations, such as serving on faculty recruitment committees or organizing lecture series. This means they have less time for research collaborations.

And unless international collaboration yields results that are valued at an institution, it will be invisible. Some institutions view international collaborations as a frivolous pursuit that's likely to count less towards tenure than are other endeavours. Everyone should ▶

check to what degree global engagement might or might not factor into promotions. Still, women should think of international research collaborations as a way of expanding their academic circles and doing the kind of internationally recognized science that gets published in high-profile journals — and that will help to secure tenure.

Women who mainly teach should seek institutional funding so that they can travel abroad with students to conduct research. They can identify potential international collaborators by searching through journals, associations and government and organizational reports to find people who do similar research. It's also a good idea to talk to international colleagues on campus or at neighbouring institutions about their research plans, in case there are overlapping or complementary interests.

I also encourage men to be allies, to make sure that women get invited to conferences and to ensure a safe, respectful environment at such events. If we wait for women alone to change the world, we'll be waiting a long time.

RICHARD DE GRIJS
Identify potential pitfalls

Astrophysicist and associate dean for global engagement at Macquarie University, Sydney, Australia.

At any given time, I am involved in five to ten collaborations. Some might have only one or two people. Most astronomy research, including mine, is driven by small collaborations. It's important that members of a team get on, otherwise the team will go nowhere. Often, I'll meet someone at a conference and we'll end up working together. Alternatively, I'll need someone else's expertise, or they'll need mine.

Multinational collaborations, which often require government investment, come about in different ways. But as long as the collaborators respect each other scientifically, things should be fine. As I wrote in 'Ten simple rules for establishing international research collaborations' (R. de Grijs *PLoS Comput. Biol.* **11**, e1004311; 2015), you want collaborators who are dependable, reliably meet deadlines and have a good reputation. Working with such people is the best way to protect your own reputation.

I belong to a collaboration of 20–30 members who are part of an observational campaign at the European Southern Observatory in Chile. I was invited to join by a personal contact. The team has been in place for around 9 years and has published more than 30 papers.

It's essential that all the members of a team meet in person, ideally at least once every two or so years — otherwise, connections are watered down and the team becomes less effective. Meetings can also help to address any cultural differences. I'm a Dutch scientist who has collaborated extensively in the United States, the United Kingdom and China. Students in some nations might look up to you and treat you as the person who knows everything, whereas in other nations, researchers are used to being questioned. And deadlines, particularly if not strict, can be taken more loosely in some cultures than in others. In larger collaborations, and in fields such as astrophysics, which have proprietary information, it can help to have a formal agreement about what's expected of the team and how data will be used.

JOSÉ ANTONIO CASTILLO MORALES
Reach out

Biologist, Yachay Tech University, Urcuquí, Ecuador.

About 80% of publications in Ecuador result from international collaboration. For small Central and South American countries such as Ecuador and Paraguay, international collaboration is extremely important.

Few native Ecuadorean scientists have doctorates and do research, but the number of people with PhDs is rising as the national government invests in sending young people abroad for their degrees. I work at Yachay Tech University, which opened in 2014. Roughly 80% of the faculty members there are from other countries. We are funding fellowship programmes to train more PhD students in hard sciences. We require students to learn English, in part because that is the language of science publications. Students are wise to build solid collaborations with their advisers overseas, and to keep those going when they return. Ecuador has cultivated strong ties with programmes in Belgium, Spain and the United States.

That said, most of our international collaborations begin because a foreign partner is seeking one, particularly in areas related to biodiversity, agriculture and medicine. Ecuador is rich in plants and animals, and continues to do research on Zika, malaria and dengue.

Researchers at Ecuador's big universities, particularly English speakers, are the most likely to be interested in joining a collaboration. But, because many Ecuadoreans do not speak English, it might be best to send them a message in Spanish, for example using Google Translate, to increase your chances of a response.

By law, students who have fellowships or studentships funded by the national government are required to come back to Ecuador, either to teach or to help the country in some other way using their knowledge. The government is trying to open posts in universities and companies to attract these newly qualified master's and PhD holders.

LEIF OLTEDAL
Formalize agreements

Neuroscientist, Global ECT-MRI Research Collaboration, University of Bergen, Norway.

In 2013, while using magnetic resonance imaging (MRI) to study neurological responses to electroconvulsive therapy (ECT) as a treatment for depression, I realized I was often being scooped by other groups doing the same thing. Our research group in Bergen established a collaboration with Anders Dale, a neuroscientist at the University of California, San Diego. Next, I searched PubMed to identify all the other groups that were using radiology before and after ECT. Anders agreed that we should invite other groups to join a collaboration to analyse data longitudinally.

In 2014, I contacted 12 groups, asking them to collaborate. Together, we could analyse data from 150–200 patients instead of the more typical 10–20. Twenty minutes after I sent the first e-mail, I got a reply from one group saying that it was a great idea. More people answered over the next few days. We were able to invite everyone — 13 international participants from 10 sites — to Bergen for a 2-day meeting in June 2015. I was surprised by the number of groups that expressed interest. Our collaboration with Anders was crucial to the response. I'm not sure that we would have got such numbers had the invitation come just from me, then a postdoc in radiology.

Before our collaboration's first meeting in 2015, we had four group conversations in which people introduced themselves and spoke about the research projects they were working on. We established a pact for the collaboration that included rules on authorship and our data-sharing agreement. We have not had direct conflicts, but we have discussed how to manage potential conflicts should they arise.

For example, we created a shared repository for project data on a common server. Any collaborator who has contributed data can suggest what kinds of analyses he or she wants to do. If the team agrees, the person can go ahead and do the analyses. We decided that collaborators retain the right to their own data. Therefore, if one of us wants to do an analysis, anyone who doesn't think it's a good idea can request that their data not be included. We have been able to resolve any issues arising when two groups want to do the same analysis — for example, we discuss whether the groups can collaborate, or do different analyses sequentially. ■

INTERVIEWS BY VIRGINIA GEWIN

interviews have been edited for length and clarity.

THE PREPRINT

Time to publish.

BY J. W. ARMSTRONG

ILLUSTRATION BY JACEY

It is not widely known that there's a machine at the centre of the Universe, the sole function of which is to create more time.

Now, a physicist might say 'time' is the independent variable in the equations of motion. A poet might wax lyrical about how precious it is to young lovers. Perhaps a mystic would speak of things unknowable. Most of us would say it's the stuff that moves too slowly when we're young and too quickly when we're old. It's the ultimate asset: the thing, especially near the end, that everyone wants more of.

In any event, it's consumed, well, all the time. Hence the necessity of the machine, to continually produce more of it. The machine has been around from the beginning, faithfully creating all the time the Universe needs. It will continue to do so for the rest of eternity.

I inferred the existence of the machine in (to be both immodest and truthful) a stunning display of mathematical virtuosity. I was investigating aspects of the standard model when I hit upon these beautiful equations. I was thunderstruck by their *necessity* — and their implications. This was the biggest discovery in human history and, of course, I *had* to get proper credit for it.

The obvious thing was to publish quickly. After all, publication and priority are how academics keep score. But, following custom, I first gave a seminar at the university. My colleagues' reaction: "A machine? That creates time?" They left shaking their heads, confident I had gone bonkers. None could see the machine's existence followed *inevitably* from the mathematics. With wounded pride (and diminished regard for my colleagues), I withdrew to my lab.

I studied the equations. I realized they implied a device could be built for transport — first me, then my doubters — to see the machine. The ultimate proof! Construction proved difficult but feasible. Details are in my lab notebooks. For reasons that will become clear later, my preprint on this topic will not be immediately forthcoming.

When I finished the device — and with only minor trepidation — I energized it and was immediately *there*.

The machine is, of course, sentient. It also does not like visitors. Or even the idea that anyone knows it exists. When I arrived, I was not welcome.

I wondered how the machine might communicate. Telepathically? Via avatar? Anticlimactically, it was just in a disembodied conversational voice.

Or perhaps telepathy *was* involved, because it answered my initial question before I asked it. "You're not the first to infer my existence," the machine commented sourly. "But you are the first to devise a way to find me. Previously only mystics could divine my being. They explained me as an elemental property of the Universe. But, lacking proof, they were judged to be mad. I was safely forgotten for a time."

It paused. "You'd have got better acceptance from your colleagues describing me as a quantum field — 'an elemental property of the Universe' — rather than a clockwork mechanism."

Maybe it had a point. But I didn't like being lectured, even by an entity 13.8 billion years old. Fascination with unfolding events kept me silent, however.

The machine sighed. "Being human, I assume you seek some boon. Immortality? Transport back in time to relive your youth? Even if I were so inclined — which I'm not — such actions are on The List, prohibited except in extraordinary circumstances."

I objected that I had come with purity of motive, just to understand how the Universe works. The machine seemed sceptical. "Really? Well, good for you." It paused. "Since you're solely interested in knowledge, I'll answer the obvious questions. Is the Universe causal? Approximately. Are events predetermined? Partially. Do humans have free will?" It chuckled, as if recalling a private joke. "Sure, why not?

"But to the current problem: your arrival here is unfortunate. I can't simply return you, pretending this never happened. You academics can't control yourselves. You'd publish. Others would read your paper, build transport devices and, ugh, there would be *tourists*."

It paused thoughtfully. "Of course, the obvious solution is to terminate your existence now…"

I didn't like where this was going, but before I could object the machine continued. "However, I am not without sympathy for the plight of ephemeral beings." It cleared its throat and began lecturing me again. "Knowledge and the ability to act on knowledge are different. Your transport device required sophisticated technology. And without the ability to travel here — to prove I exist — you're just a crazy scientist with an untestable idea. Your transport capability, therefore, advances the problem to an 'extraordinary circumstance'. I'll resolve this by sending you back in time to a pre-technical era. There you can live out your days, jabber about your equations, and be hopelessly unable to do anything else."

It paused. "Perhaps you'll even tell tales of me. None of which, of course, will be believed."

I objected that this wasn't fair! The machine replied mildly, "Ah. Fairness." It let that sink in and then added, "You should be flattered — your discovery is important enough to justify backward-going time travel to keep it a secret!" It then muttered, "Of course, *I'll* be doing paperwork forever to rationalize the required causality violation." Recovering, it continued briskly. "Do you have a preference for your exile? Neanderthal Europe 50,000 years ago? China 3,000 years ago? Carthage 2,500 years ago?"

I sighed, resigned to learning Phoenician. As the machine prepared to dispatch me, however, I was still thinking about how I might get this published. A 2,500-year lead time *does* emphasize the 'pre' in preprint. Still, if I could get the equations inscribed on stone tablets… ■

J. W. Armstrong *works at a large laboratory in Southern California.*